Digitized by Arya Samaj Foundation Chennai and eGangotri

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

J. F. N. L.

VOL 6

1989

G. K. V.
HARDWAR

CC-0. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

110144

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

110144
Digitized by Arya Samaj Foundation Chennai and eGangotri
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

110144

VOLUME LXXXVI, NUMBER 1
JANUARY 1989

page

5 *Why Dialogue?* Bruce Ackerman

23 *On Degrees* Rayme E. Engel

COMMENTS AND CRITICISM

38 *Perceptual Presentations and Biological Function: A Comment on Matthen* Alexander Rosenberg

BOOK REVIEWS

44 John Cottingham: *Descartes* Lilli Alanen

50 NOTES AND NEWS

54 MEMO TO AUTHORS

55 JOHNSONIAN PRIZE IN PHILOSOPHY

56 STATEMENT OF COPYRIGHT POLICY

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY
FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Editors: Bernard Berofsky, Arthur C. Danto, Hidé Ishiguro, Isaac Levi, Mary Mothersill, Charles D. Parsons, Gisela Striker, James J. Walsh. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Charles Larmore. *Managing Editor:* Michael Kelly. *Editorial & Business offices:* 709 Philosophy Hall, Columbia University, New York, New York 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, $25.00; institutions, $40.00; students and the retired and unemployed, $20.00; plus $7 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; Leigh S. Cauman; and André Schiffrin.

NOTE TO CONTRIBUTORS

As a monthly periodical, THE JOURNAL OF PHILOSOPHY encourages philosophical exchange and progress through prompt publication of articles of current interest. The editors now feel that they can ensure publication within six to twelve months of acceptance. They continue to be particularly interested in papers that explore the borderline between philosophy and the special disciplines, but are unlikely to accept articles primarily expository or historical in character. For the Comments and Criticism section they accept rejoinders but prefer those designed to advance the discussion rather than just to set the record straight. If and only if articles are held five months or more and then rejected, letters of rejection include a brief comment concerning the editors' reasons for rejection.

Attention is drawn to the memo to authors and to the statement of copyright policy, republished on pages 54 and 56 of this issue; copies are available from the JOURNAL office.

Contributors are supplied free of charge with ten copies of the JOURNAL issues in which their articles appear, reviewers and authors of abstracts with two copies. Offprints may be ordered from the printer on a form supplied by the printer and enclosed with galley proof; they will be sent out within six weeks of publication.

© Copyright 1989 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to *Journal of Philosophy*, 709 Philosophy Hall, Columbia U, NYC 10027.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

पुस्तकालय
233189

VOLUME LXXXVI, NUMBER 2
FEBRUARY 1989

page

57 *Realism without Internalism: A Critique of Searle on Intentionality* Akeel Bilgrami

73 *A Preference for Sense and Reference* Gabriel Segal

90 *A Paradoxical Departure from Consequentialism* George W. Harris

103 NEW BOOKS

112 ERRATA

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Editors: Bernard Berofsky, Arthur C. Danto, Hidé Ishiguro, Isaac Levi, Mary Mothersill, Charles D. Parsons, Gisela Striker, James J. Walsh. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Charles Larmore. *Managing Editor:* Michael Kelly. *Editorial & Business offices:* 709 Philosophy Hall, Columbia University, New York, New York 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, $25.00; institutions, $40.00; students and the retired and unemployed, $20.00; plus $7 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; Leigh S. Cauman; and André Schiffrin.

Social Theory and Practice

An International and Interdisciplinary Journal of Social Philosophy

Volume 14, Number 3 Fall 1988

Special Issue:

Marxism-Feminism: Powers of Theory/Theories of Power

Guest Editors: *Roger S. Gottlieb and Nancy Holmstrom*

Charles W. Mills
Harry Brod
Patricia S. Mann
Thomas E. Wartenberg
Richard Schmitt
Steven Vogel
Raymond A. Belliotti

Individuals $12
Institutions $27

Department of Philosophy, Florida State University
Tallahassee, Florida 32306

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL'S terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1989 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY and additional mailing offices. Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVI, NUMBER 3
MARCH 1989

page		
113	*Quine's Point of View*	Miriam Solomon
	COMMENTS AND CRITICISM	
137	*Johnston on Human Beings*	David S. Oderberg
	BOOK REVIEWS	
141	Gareth Evans: *Collected Papers*	Scott Soames
156	Gareth Evans: *Collected Papers*	John Campbell
164	NEW BOOKS: Anthologies	
167	NEW BOOKS: Translations	

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Editors: Bernard Berofsky, Arthur C. Danto, Hidé Ishiguro, Isaac Levi, Mary Mothersill, Charles D. Parsons, Gisela Striker, James J. Walsh. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Charles Larmore. *Managing Editor:* Michael Kelly. *Editorial & Business offices:* 709 Philosophy Hall, Columbia University, New York, New York 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, $25.00; institutions, $40.00; students and the retired and unemployed, $20.00; plus $7 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; Leigh S. Cauman; and André Schiffrin.

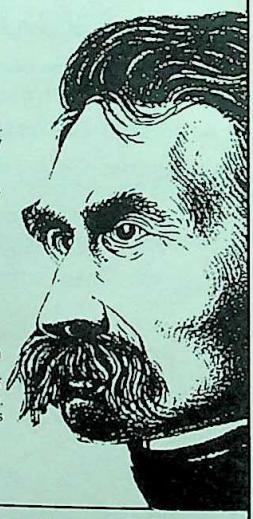
NIETZSCHE'S NEW SEAS

Explorations in Philosophy, Aesthetics, and Politics

Edited by MICHAEL ALLEN GILLESPIE *and* TRACY B. STRONG

Makes available for the first time recent Nietzsche scholarship from Germany, France, and the United States. The editors have brought together scholars from a variety of disciplines whose work is unified by a new approach to Nietzsche, one that claims that we can best understand the meaning of what Nietzsche says by coming to terms with the way he says it.

Cloth $24.95 248 pages



MORALITY WITHIN THE LIMITS OF REASON

RUSSELL HARDIN

In this provocative reconstruction of utilitarianism, Hardin brings his wide range of expertise in philosophy, economics, and political theory to bear on the subject of the practical constraints that are involved in ethical choice.

"Hardin argues forcefully and defends his general thesis well. This is an interesting and important book."
—Amartya Sen, Harvard University

Cloth $24.95 254 pages

THE UNIVERSITY OF CHICAGO PRESS *5801 South Ellis Avenue, Chicago, IL 60637*

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL'S terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1989 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVI, NUMBER 4
APRIL 1989

page

169 *The Backward Induction Paradox*
Philip Pettit and Robert Sugden

183 *Why Abortion Is Immoral* Don Marquis

BOOK REVIEWS

203 Seyla Benhabib: *Critique, Norm and Utopia: A Study of the Foundations of Critical Theory* Elliot L. Jurist

208 W. D. Falk: *Ought, Reasons, and Morality*
Stephen L. Darwall

214 Paul Guyer: *Kant and the Claims of Knowledge*
Henry E. Allison

222 NOTES AND NEWS

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Editors: Bernard Berofsky, Arthur C. Danto, Hidé Ishiguro, Isaac Levi, Mary Mothersill, Charles D. Parsons, Gisela Striker, James J. Walsh. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Charles Larmore. *Managing Editor:* Michael Kelly. *Editorial & Business offices:* 709 Philosophy Hall, Columbia University, New York, New York 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, $25.00; institutions, $40.00; students and the retired and unemployed, $20.00; plus $7 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; Leigh S. Cauman; and André Schiffrin.

OXFORD

Mind *The Journal of the Mind Association*

Editor: **Simon Blackburn,** Pembroke College, Oxford

Mind is the best known British philosophical journal, and one of the most widely read and respected journals in professional philosophy.

Forthcoming and recent articles include:

The Narrow Semantics of Names **Michael McDermott**

Personal and Impersonal Identity **T L S Sprigge**

Beyond Skepticism, to the Best of Our Knowledge **Ernest Sosa**

The Difference between Feeling and Thinking **Stephen Everson**

Can Truth be Relativized to Kinds of Mind? **Leslie Stevenson**

Subscription rates for Volume 97, 1988 (4 issues):

Institutional rate: UK £18, N America US$38, Elsewhere £22
Individual rate: UK£11, N America US$25, Elsewhere £15

To order, please apply (with correct remittance) to Wini Moranville at address below, or send for a **free sample copy**. Please note that personal subscriptions are sent to private addresses only. Payment may be made by cheque or major credit card.

Oxford University Press, 200 Madison Avenue, New York, NY 10016, USA

JOURNALS

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL'S terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1989 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to *Journal of Philosophy,* 709 Philosophy Hall, Columbia U, NYC 10027.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

A 27618

VOLUME LXXXVI, NUMBER 5
MAY 1989

page

225 *On the Shared Preferences of Two Bayesian Decision Makers* — Teddy Seidenfeld, Joseph B. Kadane, and Mark J. Schervish

245 *Science and Objectivity* — Peter Kosso

258 *Brute Experience* — Peter Carruthers

BOOK REVIEWS

270 Charles Taylor: *Philosophical Papers, 2 vols.* — Michael E. Rosen

277 NEW BOOKS

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Editors: Bernard Berofsky, Arthur C. Danto, Hidé Ishiguro, Isaac Levi, Mary Mothersill, Charles D. Parsons, Gisela Striker, James J. Walsh. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Charles Larmore. *Managing Editor:* Michael Kelly. *Editorial & Business offices:* 709 Philosophy Hall, Columbia University, New York, New York 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, $25.00; institutions, $40.00; students and the retired and unemployed, $20.00; plus $7 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; Leigh S. Cauman; and André Schiffrin.

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL'S terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1989 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVI, NUMBER 6
JUNE 1989

page

281 *Biosemantics* Ruth Garrett Millikan

298 *Metapsychological Relativism and the Self* Stephen L. White

COMMENTS AND CRITICISM

324 *The Maladroitness of Epistemic TIT FOR TAT* John Woods

BOOK REVIEWS

331 Charles Larmore: *Patterns of Moral Complexity* Jeremy Waldron

333 NOTES AND NEWS

334 NEW BOOKS: Anthologies

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Editors: Bernard Berofsky, Arthur C. Danto, Hidé Ishiguro, Isaac Levi, Mary Mothersill, Charles D. Parsons, Gisela Striker, James J. Walsh. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Charles Larmore. *Managing Editor:* Michael Kelly. *Editorial & Business offices:* 709 Philosophy Hall, Columbia University, New York, New York 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, $25.00; institutions, $40.00; students and the retired and unemployed, $20.00; plus $7 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; Leigh S. Cauman; and André Schiffrin.

Teaching Philosophy

Call For Papers

Teaching Modern Philosophy

$250 Prize For The Best Article

Teaching Philosophy is devoting a special issue to Modern Philosophy and is soliciting papers on the following topics:

- The relevance of Modern Philosophy today, or of specific authors (Descartes, Leibniz, Hume, et al.).
- Teaching the Moderns in new guises and other courses.
- Novel ways to teach the concepts of Modern Philosophy.
- Novel assignments and course designs.
- Special problems in teaching Modern Philosophy, etc.

The deadline for submissions is September 1, 1989. Please send 3 copies of your manuscript to *Teaching Philosophy*, Arnold Wilson, Editor, University of Cincinnati, Cincinnati, Ohio 45221-0206, USA.

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL'S terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1989 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY and additional mailing offices. Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVI, NUMBER 7
JULY 1989

page

337 *Leibniz, Keynes, and the Rabbis on a Problem of Distributive Justice* Nicholas Rescher

353 *Cartwright and the Lying Laws of Physics* Ronald Laymon

373 *The Virtues of Ignorance* Julia Driver

BOOK REVIEWS

385 Philip Kitcher: *Vaulting Ambition: Sociobiology and the Quest for Human Nature* Richard M. Burian

391 JOHNSONIAN PRIZE ANNOUNCEMENT

392 NOTES AND NEWS

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Editors: Bernard Berofsky, Arthur C. Danto, Hidé Ishiguro, Isaac Levi, Mary Mothersill, Charles D. Parsons, Gisela Striker, James J. Walsh. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Charles Larmore. *Managing Editor:* Michael Kelly. *Editorial & Business offices:* 709 Philosophy Hall, Columbia University, New York, New York 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, $25.00; institutions, $40.00; students and the retired and unemployed, $20.00; plus $7 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; Leigh S. Cauman; and André Schiffrin.

Call for Papers

Moral Philosophy in the Public Domain

An International Conference on Applied Ethics in Business, Medicine and Environmental Policy

June 7, 8, 9, 1990

Vancouver, British Columbia, Canada

You are invited to submit completed papers or abstracts (approx. 750 words) to:

Linda Darling, Conference Coordinator
Philosophy Department
University of British Columbia
Vancouver, B.C. Canada V6T 1W5

Deadline for submissions: November 1, 1989

Sponsored by
University of British Columbia

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL'S terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1989 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

A 21918

VOLUME LXXXVI, NUMBER 8
AUGUST 1989

page

393 *Realism and the Historicity of Knowledge* Paul Feyerabend

407 *Computation and Consciousness* Tim Maudlin

BOOK REVIEWS

433 David Lewis: *Philosophical Papers, volume II*
Frank Jackson

437 Alasdair MacIntyre: *Whose Justice? Which Rationality?*
Charles Larmore

442 Paul A. Roth: *Meaning and Method in the Social Sciences*
Graham Macdonald

446 NEW BOOKS: TRANSLATIONS

447 NOTES AND NEWS

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Editors: Bernard Berofsky, Arthur C. Danto, Hidé Ishiguro, Isaac Levi, Mary Mothersill, Charles D. Parsons, Gisela Striker, James J. Walsh. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Charles Larmore. *Managing Editor:* Michael Kelly. *Editorial & Business offices:* 709 Philosophy Hall, Columbia University, New York, New York 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, $25.00; institutions, $40.00; students and the retired and unemployed, $20.00; plus $7 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; Leigh S. Cauman; and André Schiffrin.

GRAZER PHILOSOPHISCHE STUDIEN

BAND 33/34 1989 VOLUME 33/34

WITTGENSTEIN IN FOCUS – IM BRENNPUNKT: WITTGENSTEIN

Edited by Rudolf HALLER and Brian McGUINNESS

Als Erstveröffentlichung *The first time in print*

21 Briefe von Gottlob FREGE an Ludwig WITTGENSTEIN

Aufsätze von *Articles by*

Johannes BRANDL
Rosaria EGIDI
Guido FRONGIA
Aldo GARGANI
Rudolf HALLER
S. Stephen HILMY
Hidé ISHIGURO
John McDOWELL
Brian McGUINNESS
J. Christoph NYIRI
David PEARS
Colin RADFORD
Joachim SCHULTE
Hans SLUGA
Antonia SOULEZ
Ernst K. SPECHT
Avrum STROLL
Kerstin STÜSSEL
Eddy M. ZEMACH

Herausgeber *Editor*

Prof. Dr. Rudolf Haller, Institut für Philosophie, Universität Graz, Heinrichstraße 26, A-8010 Graz, Österreich/Austria

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL'S terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1989 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to *Journal of Philosophy*, 709 Philosophy Hall, Columbia U, NYC 10027.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

पुस्तकालय
गु० कां० विश्वविद्यालय
हरिद्वार
1/12/09

VOLUME LXXXVI, NUMBER 9
SEPTEMBER 1989

page

449 *The Application of Mathematics to Natural Science*
Mark Steiner

481 *Another Argument against Vague Objects*
Francis Jeffry Pelletier

BOOK REVIEWS

493 Robert M. Gordon: *The Structure of Emotions;* and Ronald de Sousa: *The Rationality of Emotions* Marcia Cavell

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Editors: Bernard Berofsky, Arthur C. Danto, Charles Larmore, Isaac Levi, Mary Mothersill, Charles D. Parsons, James J. Walsh. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine. *Managing Editor:* Michael Kelly. *Editorial & Business offices:* 709 Philosophy Hall, Columbia University, NY, NY 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, $25.00; institutions, $40.00; students and the retired and unemployed, $20.00; plus $7 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; Leigh S. Cauman; and André Schiffrin.

THE JOURNAL OF PHILOSOPHY

Purpose: To publish distinguished philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and the special disciplines.

SUBSCRIPTIONS

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Please inquire for price lists, shipping charges, and discounts on back orders.

709 Philosophy Hall, Columbia University, NYC 10027

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL'S terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1989 by the Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027.

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVI, NUMBER 10
OCTOBER 1989

Special Issue, Eighty-sixth Annual Meeting
American Philosophical Association, Eastern Division

page

Symposium: Thought of Grice
505 *The Metaphysical Construction of Value* — Judith Baker
514 *On Grice on Language* — Richard E. Grandy
526 *On Grandy on Grice* — Robert Stalnaker
528 *Reply to Baker and Grandy* — Richard Warner

Symposium: Kripke's Theory of Truth
530 *Applying Kripke's Theory of Truth* — Vann McGee
539 *Truth, Definite Truth, and Paradox* — Stephen Yablo

Symposium: Egoism and Morality
542 *Egoism and Morality* — Leslie Mulholland
551 *Doing Justice to Egoism* — Laurence Thomas

Symposium: Philosophy of Mathematics
553 *Domain Extension and the Philosophy of Mathematics* — Kenneth Manders
563 *Innovation and Understanding in Mathematics* — Philip Kitcher

565 Johnsonian Prize Announcement
566 NOTES AND NEWS

567 Program of the Meetings
579 List of Program Participants
584 Appendix: Programs of Group Meetings
596 List of Group Participants

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Editors: Bernard Berofsky, Arthur C. Danto, Charles Larmore, Isaac Levi, Mary Mothersill, Charles D. Parsons, James J. Walsh. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine. *Managing Editor:* Michael Kelly. *Editorial & Business offices:* 709 Philosophy Hall, Columbia University, NY, NY 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, $30.00; institutions, $50.00; students and the retired and unemployed, $20.00; plus $8 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; Leigh S. Cauman; and André Schiffrin.

THE JOURNAL OF PHILOSOPHY

Purpose: To publish distinguished philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and the special disciplines.

SUBSCRIPTIONS

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Please inquire for price lists, shipping charges, and discounts on back orders.

709 Philosophy Hall, Columbia University, NYC 10027

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL'S terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1989 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVI, NUMBER 11
NOVEMBER 1989

Special Issue, Eighty-sixth Annual Meeting
American Philosophical Association, Eastern Division

page

Symposium: Thought of Michel Foucault

601 *Foucault and the Paradox of Bodily Inscriptions* — Judith Butler

608 *Culture or Nature? The Function of the Term 'Body' in the Work of Michel Foucault* — Ladelle McWhorter

615 *Foucault's Critique of the Liberal Individual* — David F. Gruber

Symposium: Thought of Wittgenstein

622 *Wittgenstein's Later Philosophy of Mind: Sensation, Privacy, and Intention* — Crispin Wright

635 *Wittgenstein, Mind, and Scientism* — Warren Goldfarb

643 *Wittgenstein and the Inner World* — John McDowell

Symposium: Impartiality

645 *The Impracticality of Impartiality* — Marilyn Friedman

657 *Friedman on Impartiality and Practicality* — Howard McGary

Symposium: Analytical Marxism

659 *How Marxism Became Analytic* — Bill Martin

667 *Reason and Explanation* — Andrew Levine

Symposium: Thought of Adam Smith

669 *The Role of Self-interest in Adam Smith's Wealth of Nations* — Patricia H. Werhane

681 *Adam Smith on Virtue and Self-interest* — Charles L. Griswold, Jr.

683 NOTES AND NEWS

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Editors: Bernard Berofsky, Arthur C. Danto, Charles Larmore, Isaac Levi, Mary Mothersill, Charles D. Parsons, James J. Walsh. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine. *Managing Editor:* Michael Kelly. *Editorial & Business offices:* 709 Philosophy Hall, Columbia University, NY, NY 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, $25.00; institutions, $40.00; students and the retired and unemployed, $20.00; plus $7 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; Leigh S. Cauman; and André Schiffrin.

THE JOURNAL OF PHILOSOPHY

Purpose: To publish distinguished philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and the special disciplines.

SUBSCRIPTIONS

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Please inquire for price lists, shipping charges, and discounts on back orders.

709 Philosophy Hall, Columbia University, NYC 10027

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL'S terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1989 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVI, NUMBER 12
DECEMBER 1989

पुस्तकालय
डी० वि० विश्वविद्यालय
हरिद्वार

page

685 *The Prince and the Phone Booth: Reporting Puzzling Beliefs* Mark Crimmins and John Perry

712 *Social Power and Human Agency* Andrew Kernohan

COMMENTS AND CRITICISM

727 *Intensionality and Perception: A Reply to Rosenberg* Mohan Matthen

734 NOTES AND NEWS

735 STATEMENT OF OWNERSHIP

736 INDEX TO VOLUME LXXXVI

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Editors: Bernard Berofsky, Arthur C. Danto, Charles Larmore, Isaac Levi, Mary Mothersill, Charles D. Parsons, James J. Walsh. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine. *Managing Editor:* Michael Kelly. *Editorial & Business offices:* 709 Philosophy Hall, Columbia University, NY, NY 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, $30.00; institutions, $50.00; students and the retired and unemployed, $20.00; plus $8 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; Leigh S. Cauman; and André Schiffrin.

THE JOURNAL OF PHILOSOPHY

Purpose: To publish distinguished philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and the special disciplines.

SUBSCRIPTIONS

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Please inquire for price lists, shipping charges, and discounts on back orders.

709 Philosophy Hall, Columbia University, NYC 10027

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL'S terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1989 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVI
JANUARY–DECEMBER 1989

Edited by
Bernard Berofsky, Arthur C. Danto
Hide Ishiguro, Isaac Levi, Sidney Morgenbesser
Mary Mothersill, Charles D. Parsons
Gisela Striker, James J. Walsh

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL'S terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1989 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to *Journal of Philosophy*, 709 Philosophy, Columbia U, NYC 10027.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVI, NO. 1, JANUARY 1989

WHY DIALOGUE?

BEGIN by considering the role of dialogue in the life of a morally reflective person—a person, that is, who seriously asks himself how he should live and tries to live his life according to the answers he finds most plausible. How does talking enter into this exercise in self-definition?

Well, it all depends. Surely Socrates's response remains relevant: compared to the question of the good life, do not all others pale by comparison? And how better to discover the truth than to engage in discussion with anyone and everyone who professes an answer?

And yet I cannot allow Socrates to monopolize my moral vision. His fixation upon his single question threatens to distort the shape of human life—when will I get on with life itself if I am forever hung up on the question of how I ought to live? This is not to say that I should close myself off from moral questioning after adolescence; only that I must walk a tightrope, taking irreversible steps in the moral twilight while retaining the courage to stop and pause later on to peer among the shadows—at the risk of learning that my earlier decisions were worthless or worse.

This reflection leads to second thoughts about the happy image of Socrates roaming the forum in search of the next know-it-all who might be snared into serious conversation. If answering Socrates's question does not amount to the whole of my life, I shall have to be far more selective about my conversation partners. I shall have to listen hard when others tell me that one book is a waste of time, another worth reading; this person a fool, another worth talking to. The need for extreme selectivity engenders, in turn, doubts about the value of the dialogues I actually do conduct. Perhaps there is some Socrates out there who would, if I only spent more time searching for him, help me discover the absurdities in the plausible–sounding–stuff–I am–listening–to?

0022-362X/89/8601/5–22 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Such anxieties may lead to the exploration of two very different paths—both of which lead away from a dialogic view of ethics. One path leads to the question of faith: Given the fragmentary character of the real-world dialogues I conduct, the real question is whom should I trust in helping me pick appropriate talking partners—which church or school or tradition?

But I need not ask myself this question. Instead of appealing to external authority, I may appeal to myself. "Bruce, don't be too impressed with what Kant said, or your Aunt Selma for that matter. You will have to think for yourself. Talking to others is no substitute. Admittedly, your own moral insight isn't much to write home about; but, in the end, it's all you've got."

For all its hubristic dangers, I subscribe to this individualistic view. For the present, though, I am not interested in defending individualism against the partisans of authority. Instead, I want to emphasize an aspect of moral phenomenology which both sides recognize. As a real-world matter, neither side supposes that talking to others is of supreme importance in moral self-definition. The key decisions are made in silence: Whom to trust? What do I really think? Although talking to others can be useful in thinking things through, a little talk may go a long way; a lot may lead nowhere. The moral value of my life does not merely depend on how I rationalize it in conversation, but upon the intrinsic value of my moral beliefs, and my success in living up to them. Thus, I do not really think less of somebody who refuses the offer of some would-be Socrates: "Stop bothering me, Socrates, I've better things to do with my time than blab with you." Such a response may well be a symptom of some deeper spiritual disease; but it may merely serve as a marker for the mature recognition that there is more to the moral life than mere talk.

All this gives me a real problem when I turn from personal morality to public life. Here I want to reverse field and proclaim dialogue as the first obligation of citizenship. Although a morally reflective person *can* permissibly cut herself off from real-world dialogue, a responsible citizen *cannot* with similar propriety cut herself off from political dialogue.

My task is not merely to suggest why this is so, but to confront the asymmetry problem: to explain why dialogue seems so much more fundamental in public than in personal life.

I.

But perhaps I am making problems for myself? Perhaps there is no fundamental asymmetry between the status of dialogue in public and personal life?

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

One way to reestablish symmetry, of course, is to demote political dialogue and deny that it is any more exigent than talk in other domains. A second way is to reassert the ultimately dialogic character of all morality—personal as well as public. I shall start by considering this second, more positive, way of reestablishing symmetry: if, despite appearances, dialogue is more generally important to morality than I have supposed, then maybe I am wrong in thinking I must solve the asymmetry problem to redeem a dialogic view of political life?

To take this line, it is not necessary to ignore the moral phenomenology with which we began. One need only deprive it of foundational importance. Sure, each of us may appropriately shut himself off from moral dialogue on countless occasions in personal life. But real-world exigencies should not blind us to the regulatory role of an ideal speech situation in our search for moral enlightenment. Despite our excusable failure to live up to the discursive ideal, we nevertheless suppose that we *should* be prepared to justify our moral choices if we ever found ourselves in an ideal dialogue with Socrates, Freud, and whomever-else-might-be-around-at-the-time. Indeed, we implicitly claim as much whenever we try to justify our life to ourselves (or others). To wax ontological: moral truth just is the name we give to those conclusions which would be reached in an ideal speech situation; nothing more and nothing less.[1]

This, transparently, is quite a mouthful; and I am skeptical about its ultimate validity. For the present, though, it is not necessary to explore these doubts with you. Even if we ultimately came to endorse the dialogic ideal of morality, I do not think we will do so in a way that reestablishes symmetry with the dialogic responsibilities of public life. To explain why, I must emphasize that I am not interested in emphasizing the role of talk in some ideal world we shall never inhabit. I am talking about the very imperfect world in which we live. So long, then, as the partisan of the ideal speech situation excuses the real-world evasion of dialogue when people confront the moral dilemmas of personal life, he is still left with a troublesome asymmetry: Why does a dialogic ideal that allows such easy escape by

[1] This theme has been developed at great length, and from different angles, by Jürgen Habermas throughout his career. Compare his *Knowledge and Human Interests,* Jeremy J. Shapiro, trans. (Boston: Beacon, 1971) with *Legitimation Crisis,* Thomas A. McCarthy, trans. (Boston: Beacon, 1975) with *The Theory of Communicative Action,* Thomas A. McCarthy, trans. (Boston: Beacon, vol. I: 1984; vol. II: 1987).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

real-world people in personal life *not* allow such an easy escape when these very same people confront the dilemmas of public life?

This is, I think, quite a hard question to answer—so long, at least, as we continue to look upon dialogue as a tool for the discovery of moral truth. For the world of practical politics does not seem at all close to anybody's idea of an ideal speech situation. Politicians do talk a lot, but it is not unduly cynical to suppose that they mean less of what they say than other folk. Thus, if I were trying to carve out a limited area of practical life where I had an especially strong obligation to participate in the dialogic search for moral truth, politics seems an unlikely place to start. However imperfect a philosophy classroom or church discussion group or psychoanalytic office may seem in comparison with the ideal speech situation, surely they are better bets than the public forum? If we do not think we are morally required to talk in one or another of these more private places, why should the search for moral truth require us to take talk seriously in politics?

The Athenians killed Socrates after all; and the modern state has hardly been more hospitable to the spirit of moral philosophy.

II.

Such reflections naturally lead me to question the asymmetry thesis from the other side. If practical political life seems such an unlikely forum for the dialogic search for moral truth, why insist that dialogue is especially central to politics?

Because there are other important things to talk about than the moral truth: in particular, how people who disagree about the moral truth might nonetheless reasonably solve their ongoing problem of living together. This is, at any rate, how liberals characteristically formulate the problem of political order;[2] and it is from this vantage point that I wish to vindicate the asymmetry thesis. That is, I do not propose to base my case for public dialogue on some assertedly general feature of the moral life, but upon the distinctive way liberals conceive of the problem of public order. This means, of course, that my argument will not convince people who reject the underlying liberal problematic. The most I can do here is to set up the problem in a way that makes rejection difficult.

[2] Compare T. M. Scanlon, "Contractarianism and Utilitarianism," in Amartya Sen and Bernard Williams, eds., *Utilitarianism and Beyond* (New York: Cambridge, 1982), pp. 103–128, esp. pp. 110–119; Thomas Nagel, "Moral Conflict and Political Legitimacy," *Philosophy and Public Affairs*, XVI, 3 (Summer 1987): 215–240.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Consider then a simple model of a liberal polity, consisting of *N* primary groups. Each primary group consists of one or more people who have combined faith and reason in an effort to search out the moral truth. Members of the same primary group have come up with the same answer to Socrates's question; members of different groups, different answers. Despite their ongoing disagreements, all groups find themselves on the same planet, in potential conflict over the planet's scarce resources. Hence the problem of liberal politics: How are the different groups to resolve their problem of mutual coexistence in a reasonable way?

Assume the position of *P*, a morally reflective member of one or another primary group, and you can begin to see why dialogue seems especially important for the successful solution of *this* problem. After all, if *P* could be confident that people outside her group, the not–*P*s, had reached the same conclusions as she did in their search for moral truth, then talking to them about the problem of coexistence might not be so important. Instead, each individual group could unilaterally declare its moral truths "to be self-evident" and authorize governmental officials to resolve all conflicts by consulting these self-evident truths. Since, on this hypothesis, each group's truths are identical to the others, there would be no need for the groups to talk *to one another* in order to come to terms about the aims of legitimate government. It is, however, precisely because *P* has reason to know that the not–*P*s have not reached the same moral conclusions that talk seems especially exigent. Whereas there may only be one moral truth, there are surely an infinite number of paths to moral error. Precisely because *P* believes that she has got closest to the truth, all she knows about the not–*P*s is that each has taken one of an infinite number of wrong turns on the path to truth. But which wrong turn?

The answer will be critical in the solution of the liberal problem of coexistence. Perhaps the moral "errors" of the not–*P*s will prove so serious to *P* that she will be unable to work out reasonable terms of coexistence with her fellow earthlings; but perhaps not. It all depends upon the particular moral mistakes made by the not–*P*s and how these "errors" relate to the disputes that *P* and the not–*P*s must resolve before they can live together on the same planet. There is, moreover, only one way for *P* to find out how matters stand. And that is *to talk* to the not–*P*s about it.

In undertaking this exercise in liberal conversation, *P* is not to try to convince the not–*P*s to change their minds and see, at long last, the compelling truth of *P*. Instead, the conversation has a more

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

pragmatic intention. It recognizes that, for the moment at least, neither *P* nor not–*P* is going to win the moral argument to the other's satisfaction, and proceeds to consider the way they might live together despite this ongoing disagreement. *P*'s refusal to talk to the not–*P*s about this pragmatic question simply disqualifies her from the liberal project. She cannot think of herself as a participant in a *liberal* state unless she is willing to participate (in one way or another) in this ongoing conversation with the not–*P*s. In contrast, *P* *can* think of herself as a participant in the search for moral truth without ever talking about this subject to the not–*P*s (or to her fellow *P*s for that matter).

It is this simple difference which motivates the asymmetry thesis: the liberal citizen must recognize a dialogic obligation of a categorically different, and more imperative, kind than he does in his personal pursuit of the moral truth. To put my point paradoxically, it is precisely because the liberal state does *not* aim for moral truth that its citizens must recognize themselves under such peremptory dialogic obligations. Let us call this the *supreme pragmatic imperative:* If you and I disagree about the moral truth, the only way we stand half a chance of solving our problems in coexistence in a way both of us find reasonable is by talking to one another about them.

III.

But is the pragmatic imperative really supreme? Granted, dialogue may be one way of reasonably solving the liberal's problem of coexistence; but is it the only way? If not, why should we prefer dialogue over its competitors?

These questions motivate the centuries-long liberal fascination with the marketplace as an alternative to more dialogic forms of dispute resolution. Here I am, needing some bricks to complete my cathedral; there you are, needing some beer to complete a wonderful Sunday afternoon at the football game. Do we really have to talk to one another to resolve our problem in mutual coexistence? Why do I not just trade some of my beer for some of your bricks at a mutually satisfactory rate of exchange? Why is trading not a perfectly reasonable way of side-stepping our moral disagreement over the relative merits of our Sunday activities? What does dialogue have to do with it?

My answer is that dialogue has a lot to do with it, and it is question-begging to pretend otherwise. The point is hardly novel, but bears repeating, given the ease with which modern philosophical marketeers manage to evade it. It is the old question of *meum* and *tuum*. Of course, once I agree that those bricks over there are rightfully called "yours," and you agree that this beer over here is right-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

fully "mine," we may then side-step our moral disagreements by trading away to our hearts' content.[3] But why does the marketeer suppose that both of us are so ready to indulge his complacent assumptions? After all, the question of distributive justice has not just arisen in the history of Western civilization.

Moreover, something special happens to the marketeer's pretensions to sweet reasonableness once I respond to his eager offer to sell me some bricks for my cathedral by denying that the bricks are rightfully his to sell in the first place. As I move toward the bricks with appropriative intent, our marketeer has but two choices. One is to use brute force to repel my assertion of superior right; the other is to engage me in a conversation that seeks to persuade me that, despite our disagreement over the ultimate value of cathedrals and football stadiums, it is reasonable for me to recognize the legitimacy of his superior claim to the bricks. In making this claim, the "brick-owner" may try to link up his assertion of right to the liberal virtues of a market system. Moreover, depending on the existing distribution of property rights, he might well convince me.

But this in no way defeats my point. We are not now trying to locate the place of the market within a discursive theory of political justification. We are considering whether market forms of coordination can plausibly allow liberals to deny dialogue the fundamental place accorded it by the supreme pragmatic imperative. The answer is "no" so long as the marketeer is prepared to concede that we may appropriately question each others' entitlements to the bargaining chips we bring to the bargaining table.

This seems worth emphasizing at a time when some of the best known free-market tracts ostentatiously fail to satisfy the pragmatic imperative's demand for dialogic legitimation. Robert Nozick, for example, does not deny that a satisfactory defense of market relations requires a theory of justice which defines the conditions under which one person might rightfully appropriate something his competitors also desire. Remarkably enough, he does not even try to come up with such a theory in his well-known *Anarchy, State, and Utopia,*[4] charmingly suggesting that it would be a mistake to hold off publication until perfection were reached (*ibid.*, pp. xiii–xiv).

Perhaps this was good enough for a young man writing in the early 1970s; fifteen years later, I begin to grow suspicious about his continuing silence: Does Nozick simply suppose the question of distributive justice will go away because he refuses to answer it? Why

[3] So long, at least, as our trades do not adversely affect others (a big proviso).
[4] New York: Basic Books, 1974, cf. p. 150.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

should others take his defense of markets seriously when he does not even try to answer the most obvious question about them?[5]

IV.

But how to discharge the dialogic burden of liberalism? How to talk to people who disagree with you about the moral truth? One thing is clear. Somehow or other, citizens of a liberal state must learn to talk to one another in a way that enables each of them to avoid condemning their own personal morality as evil or false. Otherwise, the conversation's pragmatic point becomes pointless.

This has hardly been lost upon the great spokesmen for the liberal tradition from Hobbes to the present. Indeed, the history of liberal thought can be read as a series of efforts to provide conversational models that would enable political participants to talk to one another in an appropriately neutral way. Rather than sustain self-confident liberal dialogue, however, these efforts have helped generate widespread skepticism about the conceptual possibility of neutrality. This skepticism poses an even greater threat to the pragmatic imperative

[5] Among modern free marketeers, only Friedrich Hayek has sought to challenge the supreme pragmatic imperative outright. Unlike Nozick, Hayek does not merely fail to answer the question of initial entitlements; he tries to cure us of the Enlightenment disease that leads us to think the question is worth asking. Against the Enlightenment, Hayek emphasizes the pathetically limited capacity of individuals to understand their environment. On his view, men and women simply are not built to make the kinds of global judgments about social conditions of the kind propounded by theorists of distributive justice, liberal or otherwise. Instead of such blundering efforts "to correct" the market, the thoughtful person should stand in awed appreciation of the way markets allow imperfect humans to exchange far more information than any one of them could possibly process on his own. Rather than destroy this delicate evolutionary organism, we should root out the Enlightenment fantasy that men and women could, through public dialogue, improve upon the invisible hand of the market. This hubristic conversation about social justice will only empower remote bureaucrats to aggrandize themselves at everybody's expense. So let us just stop asking the question of justice, and limit ourselves to bargaining with whatever bargaining chips the market is gracious enough to give us. See Hayek, *Law, Legislation, and Liberty: The Mirage of Social Justice* (Chicago: University Press, 1976), esp. pp. 62–101.

I am thoroughly unconvinced by Hayek's critique. Not that I wish to minimize the importance of markets. They are a key tool by which people with radically different ideals may coordinate their activities to mutual advantage. It is only Hayek's effort to treat the market as some quasi-divine adaptation to the human condition that I find mystifying. Rather than an organic development, the modern market system is the product of an unending series of highly self-conscious decisions by politicians, lawyers, policemen. Despite their mental limitations, most people are perfectly aware of this, and are quite capable of considering whether the market might operate more fairly if the relevant politicians, lawyers, policemen were given rather different operating instructions. The point of this footnote, however, is hardly to deal with Hayek's arguments with the care they deserve, but to point out that they proceed on a level that is more fundamental than most.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

than does a question-begging faith in the market. Whatever reasons Kant may have had for his great maxim, surely liberal pragmatists will not question the wisdom of "Ought implies can." If we cannot find a way to talk to one another neutrally, we do not seem to have much choice but to give up on the pragmatic imperative and return to the age-old effort to base political life on the truth, the whole truth, and nothing but the truth about the moral life.

To make matters grimmer, I can hardly deny that the history of liberal thought gives substance to this skeptical suggestion. Although many have sought to blaze a path to neutrality, the goal has proven disturbingly elusive. Given centuries of failure, perhaps I am naively optimistic to suppose that we should continue the quest. Naive or not, this is what I do suppose. Simple prudence, however, suggests the wisdom of studying the mistakes of the past before trying to formulate a more defensible conception of a worthwhile neutrality. Although I can hardly present a comprehensive catalog of liberal error, it will be useful to isolate three recurring false moves before trying to break out of the impasse.

The first false move tries to isolate a single value that all people consider most important despite their transparent disagreements over other values. By focusing upon the political implications of this supreme value, perhaps all of us may talk our way to a solution to our problems of coexistence in a way that we all find reasonable?

For this conversational maneuver to succeed, all groups must identify the same value as supreme. The most promising candidate, as Hobbes saw early on, is the fear of death. What is equally clear, to me at least, is that neither the fear of death nor anything else has the moral resolving power Hobbes hoped for it. As a sometime resident of New York City, I have cultivated a healthy appreciation of life's evanescent quality. Yet I am happy to report that I have not promoted self-preservation to the supreme place in my moral scheme. Indeed, I would despise a person who was not willing to sacrifice his life for more important things on suitable occasions. Thus, I would refuse to participate in a political conversation that began: "However much you and I disagree on other matters, we both accept the supreme importance of self-preservation." Rather than provide me with a neutral starting point, a Hobbesian political conversation would constantly oblige me to say things I found morally demeaning, despicable, false. If the point of liberal conversation is to enable me to talk to you without affirming moral propositions I think are false, the Hobbesian line goes nowhere.

So does a second well-worn path to neutrality. Here I am not

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

called upon to affirm the existence of a single supreme value that trumps all our lesser moral disagreements. Instead, I am invited to translate my disagreements into a specially sanitized evaluative framework that promises to purge them of their non-neutral aspect. The classic example is Jeremy Bentham's felicific calculus, based on the neutral-seeming principle that "pushpin is as good as poetry." Once citizens of a Benthamite polity learn to translate their disputes into the common denominator of utility, they are promised a discourse that will enable them to discuss their conflicts in a technocratic way that requires none of them to say anything inconsistent with their primary moral beliefs.

The trouble comes, however, when we inquire about the manual that assesses the value of each human activity in terms of its "utility." My point is not the "standard" epistemological objection—whatever it may be—to interpersonal comparisons of utility. Of course, if it proves impossible to make interpersonal comparisons, the utilitarian can hardly make good on his promised translation manual, and we have reached a dead end very quickly. Since I believe that this objection is overblown,[6] I do not want to rest my case against the translation strategy on epistemological grounds. Instead, I shall suppose that my utilitarian has emerged triumphant over his epistemological critics, producing a manual that plausibly compares the "utility" produced by pushpin, poetry, and much more besides.

It is at this stage that I want to enter my neutrality objection. However much one might admire the manual as a work in technocratic translation, it is quite another thing for the utilitarian to insist that his sanitized language must be used by the rest of us in talking about our political problems. Instead, his demand that all of us use his manual will provoke the very kind of un-neutral dialogue that he wished to avoid in the first place: Why does my desire for pushpin merit two utility points while your desire for poetry merits four? If *this* is what utility means, I refuse to speak a political language that obliges me to falsify my primary moral beliefs in such a systematic fashion!

So much, then, for two wrong turns down the path toward neutrality. Let us consolidate our ground by calling one the *trumping strategy:* here the liberal seeks to trump primary moral disagreement by positing a supreme moral value we all putatively share as the basis for

[6] For some of the reasons suggested by Donald Davidson in "Judging Interpersonal Interests," in Jon Elster and Aanund Hylland, eds., *Foundations of Social Choice Theory* (New York: Cambridge, 1986), pp. 195–211.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

neutral political dialogue. Call the second the *translation strategy:* here we are invited to translate our moral categories into some allegedly uncontroversial framework of political assessment. Since neither strategy looks at all promising, we can now give the skeptic's question about neutrality a sharper edge: Is there a liberal way of responding to primary moral disagreement without trying to trump it or to translate it?

V.

Well, I suppose we can try to transcend it. Thus Rawls, especially in his Kantian phase, seemed[7] to be inviting us to gain perspective on our primary moral disagreements by trying to strip away all the particular life experiences that make them seem so important to us. If we might only think our way to the astonishing ignorance of the inhabitants of the "Original Position," perhaps we could solve our problems in political coexistence by talking to one another in a neutral way?

Rawls's proposal is only the last in a long line of liberal exercises in transcendence. Many a utilitarian, for example, has urged us onward and upward to a perspective that, when compared to Rawls's, seems downright attractive. At least we are not obliged to think of ourselves as ignorant ciphers whose principal slogan is "more-for-me"; we are instead to talk to one another as if we were knowledgeable and benevolent ideal observers[8] of the ongoing political struggle, concerned only to maximize the group's welfare. The formidable differences between ignorant contractor and benevolent observer should not, however, blind us to a basic flaw they have in common. Partisans of both modes of transcendence seek to charge us an admission ticket, as it were, before we may participate in political conversation. We can only join the dialogue if we can manage to speak in the accents of the approved transcendent being without falsifying our primary moral commitments. If this is the price of admission, it is perfectly reasonable to refuse to pay it: "You so-called liberals say you'll only allow me to participate in political dialogue if I address our mutual problems from the vantage point of your favorite tran-

[7] Despite Rawls's subsequent disavowal of this interpretation, ["Justice as Fairness: Political not Metaphysical," *Philosophy and Public Affairs,* XIV, 3 (Summer 1985): 223–51], I do not believe that critics were simply engaged in tea-leaf reading in finding this theme (uneasily coexisting with many others) in Rawls's major works. See Michael J. Sandel, *Liberalism and the Limits of Justice* (New York: Cambridge, 1982).

[8] See Roderick Firth, "Ethical Absolutism and the Ideal Observer," *Philosophy and Phenomenological Research,* XII, 3 (March 1952): 317–345.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

scendent being. But it is precisely this affirmation of your recipe for transcendence that I find morally objectionable."

Since Rawls's construction now dominates liberal thought, this complaint currently takes the form of eloquent attacks on the deracinating conception of the self presupposed by his thought experiment. A generation ago, when liberals took ideal observing more seriously, the complaint had to do with the extraordinary kind of self–immolation required in thinking myself into a position where the personal involvements of Bruce Ackerman counted no more (if no less) than those of Joe Shmoe.[9] Although I am sympathetic to these particular complaints, it is more important to generalize them. The root of the problem is the wrongfulness involved in requiring citizens to affirm the value of *any* particular exercise in transcendence as a necessary condition for discursive participation. Any such demand will predictably require some citizens to talk about themselves in morally demeaning and falsifying ways; and it is precisely this demand that makes the pragmatic imperative pointless. Trying to transcend our moral disagreements seems, in short, no more promising than translating or trumping them. Once again, we have reached an impasse: Is there a way out?

VI.

Yes, I think there is. It is the path of conversational restraint. The basic idea is very simple. When you and I learn that we disagree about one or another dimension of the moral truth, we should not search for some common value that will trump this disagreement; nor should we try to translate it into some putatively neutral framework; nor should we seek to transcend it by talking about how some unearthly creature might resolve it. We should simply say *nothing at all* about this disagreement and put the moral ideals that divide us off the conversational agenda of the liberal state. In restraining ourselves in this way, we need not lose the chance to talk to one another about our deepest moral disagreements in countless other, more private, contexts. We simply recognize that, while these ongoing debates continue, we will gain nothing of value by falsely asserting that the political community is of one mind on deeply contested matters. Doubtless the exercise of conversational restraint will prove extremely frustrating—for it will prevent each of us from

[9] The development of this line of criticism over the last generation can be usefully approached through the work of Bernard Williams. Compare his *Ethics and the Limits of Philosophy* (Cambridge: Harvard, 1985) with J. J. C. Smart & Bernard Williams, *Utilitarianism: For and Against* (New York: Cambridge, 1973), pp. 77–155.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

justifying our political actions by appealing to many of the things we hold to be among the deepest and most revealing truths known to humanity. Nonetheless, our mutual act of conversational restraint allows all of us to win a priceless advantage: none of us will be obliged to say something in liberal conversation that seems *affirmatively false*. Having constrained the conversation in this way, we may instead use dialogue for pragmatically productive purposes: to identify normative premises all political participants find reasonable (or, at least, not unreasonable).

To refine this simple idea, begin by clarifying its intended domain of application. In calling for conversational constraint, I am not trying to stifle the voices of those who wish to challenge one or another aspect of their ongoing power relationship with others. To the contrary, a liberal polity must allow any person to raise any question she wants to if its dialogic project is to succeed: If the point of liberal politics is to come up with solutions that all participants find reasonable, how could this possibly be accomplished if citizens are not even allowed to place all their questions on the discursive agenda?

My principle of conversational restraint does not apply to the questions citizens may ask, but to the answers they may legitimately give to each others' questions:[10] whenever one citizen is confronted

[10] The simple distinction between questions and answers helps clarify a Rawlsian project with which I am very sympathetic. In his most recent work, Rawls is concerned to distinguish liberal political life from one that is based on a "mere modus vivendi." See John Rawls, "The Idea of an Overlapping Consensus," *Oxford Journal of Legal Studies*, VII, 1 (Spring 1987): 1–25, esp. pp. 10–15. For me, this dismissive label describes a regime that extends the principle of conversational restraint beyond its proper bounds to deny citizens the unrestricted right to insist on a liberal answer to any aspect of their power relationship which they wish to question by placing it on the political agenda. Consider, for example, a polity that threatens to punish anyone who tries to precipitate serious political debate about the existing distribution of property rights or gender roles. Although the political agenda is left open to questions on other matters, each person is told that she will simply have "to live with" the status quo so far as this particular dimension of power is concerned. Such a restriction on political freedom, to my mind, categorically deprives the regime of liberal legitimacy, and reduces it to the status of a "mere modus vivendi."

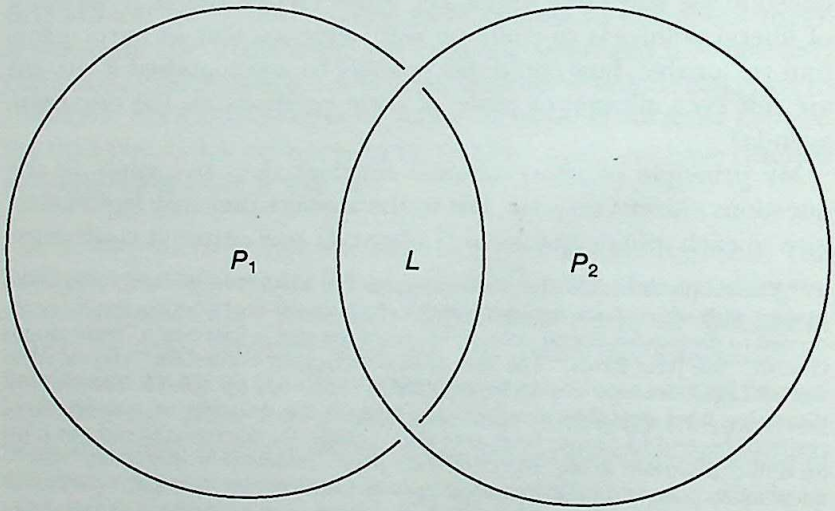
I cannot deny, however, that a "mere modus vivendi" may be the best liberals can realistically hope for under one or another extreme set of conditions—where allowing the serious political consideration of the power that comes from property or whatever will tear the place apart, and lead only to the destruction of a polity that might otherwise have generated productive political dialogue on other issues. Even as a temporary expedient, however, the use of such "gag rules" is fraught with danger. Nonetheless, the world being the place that it is, I cannot say that such a drastic step is absolutely unthinkable. See Steven Holmes, "Gag Rules or the Politics of Omission," in Jon Elster and Rune Slagstad, eds., *Constitutionalism and Democracy* (New York: Cambridge, 1988), pp. 19–58.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

by another's question, he cannot suppress the questioner, nor can he respond by appealing to (his understanding of) the moral truth; he must instead be prepared, in principle,[11] to engage in a restrained dialogic effort to locate normative premises that both sides find reasonable.

The substantive outcome of this liberal dialogue will, of course, depend upon our primary moral commitments. Nonetheless, we can say something pretty general about the formal relationship between liberal political conversation and the talk going on in other places in liberal society. To make the formal points simple, suppose that there were only two primary groups in our liberal society and that their efforts at constrained conversation have succeeded in isolating some common evaluative ground. The over-all shape of normative conversation, then, will look something like this:



P_1 and P_2 represent the set of moral propositions each primary group affirms in conversations between members of the same group. In intergroup conversation, however, participants only make use of the L-propositions for purposes of conflict resolution. For it is only these propositions which will not be condemned as false by members of either group (though the reasons P_1 and P_2 will give to justify any particular L proposition may, of course, be quite different).

While I hope you find this formalism clarifying, I should empha-

[11] Of course, practical steps must be taken to organize the agenda so as to make it manageable for real-world debate and decision. But this practical task cannot be made the pretext for the infinite deferral of a part of the agenda which some powerful participants do not wish to consider. See fn. 10.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

size that it only describes the first, and more negative, stage of the liberal dialectic—the stage where liberal citizens, through the exercise of conversational restraint, identify *L*-propositions that might function as public value premises in a liberal political argument. Even if such a purging operation were completely carried out,[12] it remains for the liberal citizenry to fashion affirmative arguments out of the available public premises—arguments sufficiently incisive to resolve the citizenry's ongoing disputes. Obviously, this affirmative operation can be a creative business. Depending on the *L*–set, there may be a host of discursive possibilities available.

For present purposes, the important thing to notice is the distinctive character of these arguments. On the one hand, none of the participants condemns any of their normative premises as morally false; on the other hand, the argumentative aim is not to discover the ultimate moral truth. Instead, it is to provide each citizen with something different—a way of reasonably responding to their continuing moral disagreement. As a consequence, the ongoing political dialogue looks very different from the kinds of conversation idealized by critical theorists like Jürgen Habermas. Most importantly, liberal citizens do not feel free to introduce any and all moral arguments into the conversational field. Instead of looking to ultimate conversational victory in some far-distant ideal speech situation, their energies are focused on the formidable task of governing *this* world through a political dialogue that does not require participants to renounce publicly their deepest moral beliefs. However humble the liberal's dialogic ambitions may seem from the unreachable heights of the ideal speech situation, my principal aim has been to convince you that even such a limited dialogic objective is not necessarily an impossible dream; that you and I may strive to govern this world through dialogue, albeit dialogue of a special kind.

VII.

But, I hear you ask, is this fourth path toward neutrality really different from the first three? Does not it too demand something from people which they may not be morally prepared to give?

Yes, but the sacrifice is of a different kind. Rather than require people to say things they believe are false, I am asking them to make a special kind of emotional sacrifice. At least on those occasions when liberal citizens meet with one another to reason their way to legally authoritative resolutions of their disputes, each must try to repress their desire to say many things which they believe are true, but which

[12] In real life, there will usually be no practical need for a sharp separation of the two phases of the liberal dialectic described in this paragraph.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

will divert the group's energy away from the elaboration of the pragmatic implications of the *L*-set.

This kind of selective repression is, I think, a familiar feature of social life. It is continually required by the ongoing exercise that sociologists call role playing. Each social role can be understood as a set of conventional constraints upon acceptable symbolic behavior. When acting as a lawyer, I operate under constraints different from those delimiting my conduct as a teacher, which are different, in turn, from those involved in working on a construction project. Some role definitions allow for a broader range of symbolic behavior than others; but all roles are constraining, placing vast domains of conversation off the agenda so long as the participants are acting within a particular role framework.[13]

Truth is not necessarily a defense for stepping out of role. Thus, when I go into a class on political philosophy, I must restrain any impulse I might have to talk about the calculus of variations—even though my remarks on the calculus may be truer than anything I have to offer about politics. To be a competent social actor, I must constantly engage in a process of selective repression—restraining the impulse to speak the truth on a vast number of role-irrelevant matters so as to get on with the particular form of life in which I am presently engaged. Just as you and I try to stick to the point when we are building a car or worshipping god, so too liberal citizens must exercise a similar kind of self-control when engaging in liberal politics—joining together neither to build a better Buick nor to save men's souls, but to solve the conflicts of social life on terms that all participants may find reasonable. Thus, in calling upon people to exercise conversational self-restraint in public life, I am asking them to exercise a fundamental competence that all socialized human beings possess (to one or another degree).

But, of course, this is hardly enough to make my demand morally uncontroversial: the idea of conversational constraint is, at best, a part of a satisfactory political philosophy, not the whole of it. Most obviously, my proposal will be opposed by the partisans of an ethic of radical spontaneity, who look upon all roles as if they were merely fetters on the human spirit, and call upon us to smash any role as soon as we begin to perform it competently.[14] On this level, the

[13] For a brilliant analysis of this aspect of role relations, see Erving Goffman, *Frame Analysis: An Essay on the Organization of Experience* (Cambridge: Harvard, 1974).

[14] Roberto Unger comes remarkably close to this position in his recent work. See his *Passion* (New York: Free Press, 1984); *Social Theory: Its Situation and Its Task* (New York: Cambridge, 1987); *False Necessity* (New York: Cambridge, 1987).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

defense of the conversational restraints of liberal citizenship is simply a special case of a more general affirmation of the value of role playing in social life. Such a defense need not deny the reality of spiritual constraint which the romantic critic finds so confining, nor the importance of working with others to create new role relations that better express the different aspects of our being. (Indeed, my proposal to develop more fully the role of liberal citizen falls precisely under this second heading.) Instead, a defender of liberal role playing should argue that the romantic critic has chosen the wrong way to ease the sense of role constriction which threatens to strangle him. Rather than assault the very idea of role playing, it seems wiser to seek relief in the marvelous human capacity to shift role engagements over time. I can be a lawyer, teacher, construction worker, father, baseball coach—as well as a liberal citizen. Although each of these roles imposes its own constraints, the value of my life is hardly exhausted by the way I confront the challenges of any particular one; it depends as well on how I shift from role to role over time and build up a meaningful whole out of these temporally-limited parts. This seems a better response to the inadequacies of any individual role than a romantic effort to repudiate role playing entirely.

Even if so much were conceded, the defense of liberal restraint must proceed to a second stage. After all, I am not arguing for any system of roles, but a system that gives the role of liberal citizenship a central place. What is so good about trying to solve disputes through a neutrally constrained dialogue?

VIII.

There are many ways to answer this question. Some will seek to elaborate the intrinsic virtues of liberal citizenship—the distinctive value of the liberal's relentless effort to weave a web of intelligibility which can link parties together despite their many moral differences from one another, the moral value of allowing all of us to participate in politics without falsifying our deepest convictions. Others will pursue more consequentialist paths—emphasizing the value of the open society that it is the aim of liberal citizenship to produce.

My aim here, however, has not been to convince you of the ultimate value of liberal citizenship. It has been to suggest its distinctive character. I have tried to explain the liberal grounds for the dialogic imperative in politics, and defend this pragmatic imperative against two relatively obvious objections—one raised by the free-marketeer, the other by the neutrality skeptic. Although the question-begging character of the free-marketeer's objection is pretty obvious,[15] I have

[15] But recall my inadequate treatment of Hayek in fn. 5.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

only begun to take the skeptic seriously. At best, I have suggested why the admitted failures of the past need not discourage us from blazing a new trail toward the liberal state—one through which, by the resolute exercise of conversational restraint, you and I may talk to one another in ways that neither of us condemns as morally unreasonable.

I have not even tried to establish affirmatively that the path of conversational restraint will not finally lead liberals to a fourth dead end. As you and I discover that we disagree about more and more things, perhaps we will find that the exercise of conversational restraint leaves us nothing to say to one another about our basic problems of coexistence. In terms of our simple Venn diagram, perhaps the *L*-set will turn out to be empty. This seems especially likely since the typical Western society contains many primary groups, whereas our simple picture only schematizes the *L*-set of a two-group society. As we increase the number of circles in our picture, the conversational space described by the intersecting *L*-set will get progressively smaller and smaller. Under modern conditions, does it shrink to zero?

This is the question I asked myself in *Social Justice in the Liberal State*.[16] Rather than rehearse my own answer, it has seemed more important to encourage you to think the question worth asking—by cautioning you against a superficial diagnosis of the liberal aspiration. Liberalism does not depend upon a question-begging faith in markets. Nor need it demand an alienating form of self-presentation which requires us to repudiate publicly our deepest moral convictions. Instead, it calls upon us to reflect upon the pragmatic imperative to talk to strangers as well as soul-mates; and to consider whether, despite the strangers' strangeness, we might still have something reasonable to say to one another about our efforts to coexist on this puzzling planet.

BRUCE ACKERMAN

Yale University

[16] New Haven: Yale, 1980.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ON DEGREES*

EVERY so often it is said that this or that is a matter of degree or a matter admitting of differences in degree.[1] For example, being warm, aware, accurate, literate, quiet, wet, and reasonable are matters that allow of differences in degree. There are, discernibly, differences in how warm, how aware, how wet, how accurate, etc., something or someone might be. For the sake of abbreviation, we may call such matters *mod* (matters of degree, matters susceptible of differences in modulation). If a property, say that of being *A*, is mod, then semantic competence requires more than the ability to say for a given thing whether it is or is not–*A*. It also requires some extent recognition, the ability to recognize differences in extent, to say how *A* something is, whether it is as *A* or not as *A* as others, or whether this or that is as *A* as it once was.

When is a matter one that admits of degree? The language is perhaps familiar but has not received excessive scrutiny. Matters most conspicuously mod are those easily viewed in a variety of degrees, e.g., being large, dark, or heavy. Wisdom, being wise or not wise, presumably admits of degree, but this is not impelled by any immediate observation. Abstractly, wisdom's *modity* (the property of being mod) can be inferred from the possibility that others were not as wise as Socrates and even Socrates was not wise. The essential clue to modity is finding a degree of *A*-ity other than merely that of being *A* and that of being non*A*. It suffices then if Meno was not as wise as Parmenides, and Parmenides was not as wise as Socrates. For modity, the exact extension of 'wise' is not pertinent, whether all three were wise, the latter two, only Socrates, or none. Crispin Wright[2] identifies the typical case of a matter of degree as one where iteration of being more and more *F* can eventuate in being *F*, or being less and less *F* can mean being not *F*. The distinctive feature is the possibility of there being more than one iteration. Possession of any property

* I am indebted to James Garson for many generous comments and productive discussions on these issues and for suggestions that helped to clarify the presentation.

[1] Notable instances include Carl Hempel's conclusion that cognitive meaningfulness and experiential significance are ultimately matters of degree, e.g., in *Aspects of Scientific Explanation* (New York: Free Press, 1965), pp. 117–133. W. V. O. Quine has also urged that the factual/conceptual distinction, the science/commonsense distinction, and epistemological footing are matters admitting of, or only of, differences in degree, e.g., in *From a Logical Point of View* (New York: Harper & Row, 1961), pp. 44–46.

[2] "On the Coherence of Vague Predicates," *Synthese*, XXX, 3/4 (April/May 1975): 325–366, p. 348.

0022-362X/89/8601/23–37
© 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

admits of at least two degrees: *A* and non *A*. The significant instances are those where exemplification is not perfectly all-or-nothing.

I. MATTERS THAT ARE MOD

Curiously, most matters, even those very mod, are generally treated as though matters of kind. The presumption may be that being mod is an exceptional condition. A closer examination suggests otherwise. Unassailable examples of matters that do not admit of degree are less abundant than might be thought. So much so, our initial thesis or hypothesis is that modity, as a property of properties, is practically ubiquitous, that is

TD: Every matter, or virtually every matter, is mod (is a matter that admits of differences in degree).

The property of being mod is itself a matter that admits of degree, is itself mod. Some matters are extensively so, some radically, and some perhaps only mootly or minutely. Some, such as biological kinds, seem paradigms of matters that indicate types rather than degrees. We have rabbits, penguins, dogs, dolphins, oaks, apples, and oranges—visible kinds. Being a cat, a frog, or a lemon might be considered more matters of kind than mod. But even being feline must have evolved in degrees. What were the parents of the earliest cats, their parents, and their grandparents? Consider the evolutionary degrees of felinity, the progressive extents in the genetically seamless ancestral chain of pre-cats and proximate cats which led up to the first cats. Even "abrupt" changes on an evolutionary scale were not literally instantaneous. "At no one point," as David Hull[3] observes, "does a structure become a lung or cease to be a lung, at no one point does one species become another" (*ibid.*, p. 52). With species, we are most familiar with discrete instances, all remarkably different from one another. We are not apt to confuse the household pets with the children or the garden vegetables. The individual species must nevertheless have evolved somewhat gradually. Also commencing within conception through to birth, there is a continuum of degrees of being a dog, rabbit, or an elephant.

Clearly, 'kindred' and kinship admit of degree; if kinds are gauged by kinship, matters of kind must be mod as well. The father of the first giraffe, though not a giraffe, was probably not remarkably different from his heir, perhaps not as aptly a giraffe, but more so than his ancestors. Imagine a definition sufficiently precise to distinguish the "first giraffe" from its father, so as to proclaim which was really the first giraffe. This would be a delicate operation. Our "first giraffe" probably resembled its parents (the last pre-giraffes and the first giraffe's closest kin) more than it resembles recent members of

[3] *Philosophy of Biological Science* (Englewood Cliffs, NJ: Prentice-Hall, 1974).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

its own species. Determinate division is thwarted by the absence of prominent natural gaps. The uniformities are not sufficiently uniform, the discontinuities are not sufficiently discontinuous. No matter how a division is made, if at all, the progression of degrees persists, and is perhaps ultimately more illuminating than searching for dichotomies, real or imaginary.

Matters most readily recognized as mod are typically those with a staggering gradient of marginal cases. Such cases are a sufficient but not a necessary condition. Being precisely cleft between *A* and not-*A* does not imply being nonmod. Regardless of how sharply the poor/not poor division is officially drawn, the different degrees of being poor or not poor are unaffected. Even where there is a natural dichotomy or litmus test for being *A*, there may still be continuously many degrees of *A*-ity. There literally is a litmus test for the matter of being acidic, but there is also a scale of many degrees of acidity and alkalinity. The unquestionableness of 'It's either acidic or it is not acidic' does not alter the degrees.

The nature of the degree structure of acidity was not an a priori determination. A literate command of the language may not suffice to judge as to what theoretical structure a given gradient might possess. Degree structures may be only very dimly and very tentatively revealed by semantic intuitions. That there would be precise theoretical boundaries or limits for 'fast', 'acidic', or 'cold' was not obvious even a posteriori, not in advance of relativity, protons, and thermodynamics. If *A*-ity is less well understood, the extent to which it comes in degrees, is cleft, metrical, or bounded may not be matters to be decided by prevailing semantic impressions.

Modity and vagueness, though sometimes confused, are very different. The degrees of acidity are no fewer, however precise the acidic/nonacidic division. Modity is not lost in the loss of vagueness. The more the degrees, the more options there are as to where division between *A* and non *A* might lie. But modity, like forking in the road, provides only the alternatives, not the irresolution. Modity itself thrives on increasing precision; it is enriched by clarity in the resolution of distinct degrees. If x and x' can at most be said to possess approximately the same degree of *A*-ity, they establish but one degree. But, if *A*-ity is sufficiently well-defined to distinguish the degree of *A*-ity exemplified by x from that exemplified by x', then at least two distinct degrees are established. Modity is thus abetted by precision, since refinements in such distinctions multiply degrees. How well degrees of *A*-ity are defined remains independent of how decisively the *A*/non *A* boundary might be drawn.

The question of how sharply a matter is cleft is itself one that admits of degree. There is no fundamental cleavage between those

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

which are precise and those which are not. The meaningfulness of a given dimension *A* does not depend on there being any sharp *A*/non *A* dichotomy; if it did, meaningfulness would not be meaningful. A matter is not incoherent if it cannot be dichotomized, nor does explanatory power hinge on there being a clear dichotomy. The hot/cold dimension provides no less in explanatory power than the acidic/alkaline dimension, though the latter is rather sharply cleft, whereas the former is not. Degrees and extents may be explanatory in ways dichotomous kinds are not, e.g., in the superiority of lawlike correlations of extents over low-grade generalizations associating kinds.

If there is no clear dichotomy between what is *A* and what is not, to this extent *A*-ity is vague or semantically defective; but this is not the final test or last word on the intelligibility of *A*-ity. There may not be any perfect unvagueness in this sense. Kant's notion of analyticity may have been ill-conceived, as W. V. O. Quine has claimed; but it is not clear how this follows from there not being any sharp cleavage between truths which are analytic and those which are not.[4] No clear line can be drawn between matters which possess such a line and those which do not. Rejection of analyticity is understandable if differences in extent are as elusive as the dichotomy, or if less elusive, not clear enough for analyticity to qualify as a profound theoretical dimension.

A matter may be recognized as mod without promising the eventual discovery of an underlying regularity or finely gradable uniformity. A term like 'analytic' may be semantically defective to the extent that its conception was not sufficiently auspicious to seize a significant dimension. For example, 'urbane' and 'schizophrenic' may not actually pick out any firm dimension. The extent to which a given term suffers such semantic defect is independent of its modity. Being urbane admits of degree, irrespective of perplexity in ordering exemplifications. Differences in extent are not well-defined in these cases. In the case of 'schizophrenic', this is evidently due to the fact that the matter remains a relative mystery; with urbanity, it may reflect a lack of keen interest in appreciably greater precision. We may not expect increasingly informative responses to questions such as: Exactly how urbane? Positive attribution is simply accepted as intelligible.

Modity does not preclude significant unrelativized attribution, e.g., asserting bluntly 'This is *A*' or 'These are not–*A*'. We might say

[4] Quine explains that his "rejection of the analyticity notion just means drawing no line between what goes into the mere understanding of the sentences of a language and what else the community sees eye-to-eye on," in *Ontological Relativity* (New York: Columbia, 1969), p. 86.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

without qualification that the clothes in the wash are wet, those in the drawer are not, and those on the line not as wet as they were. It need not be assumed that being wet means being perfectly wet, saturated, or supersaturated, or even that any ultra or ultimate degree of wetness exists.

The hypothesis TD, that virtually all matters are mod, is related, at least figuratively, to Leibniz's principle of continuity: properties must admit of varying degrees or else there could be no continuous change in passing from one to another.[5] Confirmation here, however, is largely inductive. Conspicuous modity may be rare; even if a matter is very mod, recognition of degrees may not be immediate. Matters reasonably easy to recognize as mod, however, are plentiful, e.g., being firm, just, evident, fluid, intent, brittle, secret, culpable, exact, transparent, toxic, honest, objective, flammable, and abrupt. Culpability may be shared or tempered by circumstances. Objectivity and precision may be raised or lowered. Honesty may wax with confidence, wane with anxious fear.

Notice of being flammable itself may arrest our attention without inciting questions of how much, yet some things are much more flammable than others, some much less. Immunity and munity to flame, like immunity in general, may be graded without there being any least or greatest degree. The idea that something is not flammable unless perfectly flammable may not even make good sense. Aside from being imprudent, there may not be any intelligible degree of perfect flammability. Anything not already in flame is confirmed less than perfectly flammable.

Being abrupt may seem too immediate or narrow a matter to admit of degrees. But the appearance of the dinosaurs is thought to have been less abrupt than their demise. Neither episode was as abrupt as the latest change in the weather; it less so than a change in facial expression, and that less abrupt than the disappearance of some subatomic particles. Abruptness, whatever its ostensible connotations, must admit of degrees.

Intentions are often treated as if the only question is whether intent is or is not present, as if intent were like a dial with but a single setting, having only one intensity option, full-throttle. But a person may be intent that p, intent that q, and not quite as intent that p as that q. One may be intent on going to a party, being on time, dressing in fashion, feeding the cat, mailing a letter, avoiding death, and departing before midnight—all in varying degrees.

[5] Extracting the exact nonmathematical version of the principle Leibniz had in mind is difficult. Its scope is indicated, however, by his use of the principle of continuity to predict the discovery of Zoophytes, organisms exemplifying degrees of both plant and animal.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Modity and conspicuous modity are not the same. Prominent semantic intuitions may reflect only a matter heavily laden in all-or-nothing associations, the legacy perhaps of a tacit words-are-labels semantic theory. Being rigid, for example, is often thought of in this way. But, if rigidity were not mod, there could be no wilting, stiffening, losing, or gaining rigidity. Rigidity is mod whether an utmost or perfect degree exists, whether utter rigidity is meaningful or not.

II. NECESSARY AND SUFFICIENT CONDITIONS INSUFFICIENT

The matter of being mod, or of coming in degrees, is seldom given explicit notice.[6] It is traditionally assumed that *A* is adequately defined by providing necessary and sufficient conditions, overlooking the questions of degree, extent, as *A*, and not as *A*. An exacting explanation of extension recognition is often attempted, with no explanation for our ability to recognize progression in extent, differences in the degree to which something is *A* or the extent to which it is not–*A*. Consider, for example, attempts to define disposition terms such as 'fragile'. The weakness of truth functional interpretations led, as Carl Hempel[7] recounts, to such formulations as 'x is fragile iff if *x* were to be struck sharply, then *x* would break'. Difficulties interpreting the subjunctive led to indicative analyses such as Quine's with 'x is fragile' construed as '(for some *y*) (*Mxy* and *y* breaks)' with '*M*' for 'alike in molecular structure' and the verbs understood as tenseless.[8] But note that, in whatever mood the verbs are interpreted, such analyses do nothing to accommodate important differences in degree of fragility. They do not allow the expression of any interesting lawlike relationships such as the colder *x* is, the more fragile it is. Repeatedly such notions are analysed as if each disposition were a total inclination or a total disinclination, as if such matters were not susceptible to differences in extent. Whatever benefit the regimentation of language might serve, it ought to meet a minimal test of adequacy that it not obscure such simple logical relations as the inference from '*x* is fragile, and *x* is not as fragile as *y*' to '*y* is fragile'.

Unless *A*-ity in the final analysis is accurately a simple all-or-nothing affair, then a definition providing necessary and sufficient

[6] Possibilities have been recognized as strong, small, and remote; David Lewis, for example, employs the spatial metaphor, remoteness, to give logical representation to differences in extent among "comparative possibilities" in his *Counterfactuals* (Cambridge: Harvard, 1973). That the epistemic virtues of hypotheses such as simplicity, refutability, and modesty are matters of degree is explicitly noted by Quine and Ullian in *The Web of Belief* (New York: Random House, 1978). Degrees of both negligence and neglect are treated by Michael Zimmerman in "Negligence and Moral Responsibility," *Nous*, xx, 2 (June 1986): 199–218.

[7] *Aspects of Scientific Explanation* (New York: Free Press, 1965), pp. 109–111.

[8] *Word and Object* (Cambridge: MIT, 1960), p. 225.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

conditions for being *A* is only a partial account of *A*-ity. If it is not a consequence of the definition that degrees are inadmissible, either they are admissible or the definition is incomplete. There is a clear gap in an analysis, for example, of being just or aware which omits degrees. Unless becoming just cannot be gradual, it is important whether a decision increases or decreases the extent to which the result is just or unjust. Identifying differences in degree may be at least as significant as locating a dichotomy, or defining ideal instances.

Typically, when matters are noted as admitting of degree, they are ones for which advanced means of measurement have long been in hand, for example, being warm, heavy, dark, or tall. But it should be possible for a matter to achieve philosophic discovery as mod other than in hindsight. The modity of a given matter *A* may perhaps be sensed, even in conceptual analysis, antecedent to the perfection of the standard *A*-ometer.

At some primitive level, nonmod properties, classes, and kinds form a crude conceptual scheme, attractive perhaps for its simple dichotomies, stereotypes, and compartmentalized organization, but without the potential subtlety or precision afforded by degrees. The implicit practice of treating matters of degree as sheer matters of kind may be supposed a simplifying assumption, justifiable as a convenient theoretical fiction; but it may well impede philosophical progress. It is doubtful that there is any gain in explanatory virtue or accuracy from framing theoretical principles and definitions disregarding the modity of the matters involved. Such definitions are at best incomplete, partial definitions if they fail to account for degrees. Attempts to state general principles may be more strained or puzzling than necessary. Taking degrees seriously, whatever else, ought to provide a more balanced perspective, one inviting fewer false dichotomies and untenable dualisms.

Most matters in philosophical discourse, if not readily recognized as mod, do at least lend themselves to inquiring as to differences in extent or degree. We can ask, for instance, as to differences in degree of belief, of being free or not free, of awareness, rigor, rationality, injustice, refutability, relevance, cognizance, subjectivity, certainty, indeterminacy, observationality, analyticity, clarity, goodness, fitness, personhood, preference, becoming, doubt, assent, meaningfulness, warrant, potential, virtue, value, and responsibility. In each case we can distinguish differences in extent, compare increase and decrease, and ask how much or to what degree. None is sheerly an all-or-nothing affair. It is doubtful that many are actually subject to any sharp *A*/not–*A* dichotomy or successfully captured by necessary and sufficient conditions. If we are barred from intelligible

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

questioning as to differences in extent or degree, we may be suspicious whether there is ultimately any serious sense to be made of such concepts.

III. EXTENSIONS AND EXTENTS

Certain parts of speech, notably adjectives and verbs, more readily express indications of degree, e.g., in receiving adverbial modification. Nouns seem better adapted to picking out kinds. We have learned from predicate logic not to read special import into the parts of speech, into traditional distinctions of substantive, noun, and adjective.[9] Whereas 'red' and 'canine' are adjectives, 'red thing' and 'dog' are substantives, or noun phrases. Referentially there is no difference, e.g., 'red' and 'red thing' have the same extension. The grammatical distinction may indeed be irrelevant, but, from the irrelevancy of the grammatical divisions, it does not follow that nouns and adjectives are most usefully treated as nouns or as predicates for kinds. Although traditional, it is doubtful that nouns are ideal models for thinking or theorizing about meaning. Nouns do lend themselves to the inevitable interest in reference and extension, the apparent remnants of the idea that words are labels. For a semantic correlate to 'dog', there is naturally the set of dogs, for 'red thing', the set of red things (and for 'set', the set of sets). To whatever extent they are adequate otherwise, such concepts do little to furnish differences in degree. If the semantic correlate for 'wet' is the set of wet things, the logical force of the inference 'if x is wet and y is as wet as x is, then y is wet' is unfathomable (a formal fallacy: Wx, Ayx, ergo Wy).

Extensions are inadequate as correlates even for count nouns. Consider again the progressive degrees in the ancestral chain for 'giraffe'. Extent recognition is clear but we are unable to perceive a semantic split setting off the extension; we sense the progressive extents but not the dichotomy assumed to mark the extension. A definite extension for 'giraffe' accounts for semantic competence we do not possess (ability to determine the split) but not at all for semantic competence we do possess (recognizing progression in extent). Ability to recognize progressive extents indicates semantic understanding whether or not reference or extension can be determined.

Something is missing. In viewing properties as class concepts, an important "dimension of meaning" is needlessly obscured. Modity, whether virtually ubiquitous or not, eludes the semantics of reference and extension, concepts not entirely adequate even for simple kinds.

[9] For this point of view, see Bertrand Russell, e.g., in *Principles of Mathematics* (New York: Norton, 1938), pp. 42–52. See also Quine, *Word and Object*, p. 96.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

IV. DEGREE STRUCTURES

The notion of a term's extension is limited, not only in revealing intension, as a test for sameness in meaning, but in providing for degrees, aptness, and extents. To this end, the concept of a *degree structure* characterizing the intended dimension may prove a convenient and suitably concrete addition. For a given matter, say that of being A, the degree structure for A is composed of the extension, the degrees, the extents, and the orderings they impose. The seemingly informal notion of degree can be made quite explicit. Degrees are the equivalence classes of those things equally A. The degree $[x]_A$ is the equivalence class $\{y{:}x \text{ is as } A \text{ as } y, \text{ and } y \text{ is as } A \text{ as } x\}$ of those things not different from x with respect to A-ity. (Note that 'as A as' means 'at least as A as' despite the apparent syntactical symmetry.) The ostensibly abstract talk of possessing the same degree actually has a literal first–order interpretation: $[x]_A = [x']_A$ iff x is as A as x', and vice versa. Extents and degrees are closely related. The extent of A-ity of x, $|x|_A$ is the class $\{y{:}x \text{ is as } A \text{ as } y\}$ of those things which x is at least as A as. The degree structure includes the extension and the indicated ordering of degrees and extents: $|x|_A \leq |y|_A$ iff y is as A as x. As such, degree structures may be thought of as extent structures, and thus at least figuratively as latent in the concept of extension.

Reference to degrees does not require special conventions, numbers, scales, or *A-ometers*. A fuzzy logic or multivalued approach, as advocated by L. Zadeh,[10] J. Goguen,[11] and J. L. King,[12] could be adapted to represent degrees via numerical assignments of decimal truth values. Briefly, $[x]_A$ would be the number assumed to be the truth value of 'x is A'. One difficulty, however, is the assumption that we could actually determine which exact decimal a sentence should be assigned. It is noteworthy that the precise assignments of the multivalued approaches are admitted to be arbitrary, while the ordering is what is treated as of chief significance.[13] And an ordering does not require numbers or measure functions. Degrees and extents of A-ity, in a straightforward way, are already ordered without adopting the artificial precision of sentences with decimal truth values.

The 'as A as' relation is a preordering (not a partial ordering since not antisymmetric). Degrees and extents, however, are partially or-

[10] "Fuzzy Logic and Approximate Reasoning," *Synthese*, XXX, 3/4 (April/May 1975): 407–428.

[11] "The Logic of Inexact Concepts," *Synthese*, XIX, 3/4 (April 1969): 325–373.

[12] "Bivalence and the Sorites Paradox," *American Philosophical Quarterly*, XVI, 1 (January 1979): 17–25.

[13] See, for example, K. J. Machina, "Truth, Belief, and Vagueness," *Journal of Philosophical Logic*, V, 1 (February 1976): 47–78, p. 61.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

dered, and often densely ordered as well. Between any two degrees there is apt to be a third. Invariably, if x is A and $|x|_A \leq |y|_A$, then y is A. Similarly, if x is A but y is not, then the degree of A-ity of x exceeds that of y. The degree structure for A-ity is the inverse of that for un$A$–ity: x is as A as y iff y is as un$A$ as x. Thus, x is larger than y if y is smaller than x. Note in general that neither 'x is as A as y' nor 'x is not as A as y' implies that x is A or that y is A. But, if x is not as A as y is, then y is no less A than x. Other features will vary from structure to structure.[14]

No greatest or perfect degree of A-ity may exist or make sense. There may not be any utmost degree of being distant, deep, irrelevant, large, disordered, pliable, flammable, or amusing. The notion of being perfectly flat, though lacking a physical realization, does have a geometrical interpretation, that of being utterly unflat does not, and that of being perfectly fat does not. There is no largest number, yet how fast one thing can move with respect to another is allegedly upper–bounded, theoretically by the speed of light in a vacuum.

Although reasonably understood in context, such notions as being half A, perfectly A, twice as A, ideally A, zero A, or 100 per cent A need not be taken seriously as literal. It is not without a theoretically successful investigation that the metric (if any), division, or bounds (if any) of a degree structure can be determined.

V. CONCOMITANCE

What is it for two notions to possess the same degree structure? For A and B to have the same degree structure, the terms must be at least coextensive, and also concomitant, that is, x is as A as y if and only if x is as B as y. As defined, concomitance entails identity of degrees, an identical ordering of degrees, and identity of extents: $|x|_A = |x|_B$, for all x. Any change in the extent to which something has the one property must be continuous with the change in the other. Any

[14] A systematic treatment of the logic is suggested given two axioms and a general rule of inference.

Axiom I: $[x \text{ as } Q \text{ as } y] \rightarrow (Qy \rightarrow Qx)$

Rule: $$\frac{\wedge_i^n (Qx_i \rightarrow Qy_i) \vdash (Qx \rightarrow Qy)}{\wedge_i^n [y_i \text{ as } Q \text{ as } x_i] \vdash [y \text{ as } Q \text{ as } x]}$$

Axiom II: $-[x \text{ as } Q \text{ as } y] \rightarrow [y \text{ as } Q \text{ as } x]$

These condense a number of logical principles useful in relating degrees and extents. Transitivity and reflexivity, for example, follow from the inference rule. Natural language indications of degree are more elaborate than might be thought. To accommodate being closer, for example, with 'Cab' representing 'a is close to b', then 'x as Q as y' above has to be understood as having instances 'a, b as C as c, d' representing 'a is as close to b as c is to d'.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

waxing, waning, movement, or shift in the one must be identical in detail to that in the other.

As has long been noted, being coextensive is not sufficient for synonymy, sameness of meaning, or individuating properties. Matching degree structures is a much more stringent test for matching properties and meanings. Possibly two distinct properties may by coincidence come to have the same extension, but, if they also vary concomitantly, degree by degree, extent for extent, accident is not an explanation. Identity of continuously many degrees considerably narrows the field of plausible explanations. Vast, fine–grained coincidence of intricate detail suggests that properties, if not identical, are inherently connected.

Concomitance may be used to distinguish decisively modity from vagueness. As noted, they are different dimensions. But, if modity is virtually ubiquitous, then perhaps modity and vagueness are coextensive. Since modity is either unaffected or enhanced by decreasing vagueness, the two are not concomitant.

Concomitance provides a test for properties in general, first–order, second–order, extensional, or intensional. Consider an intensional property such as awareness, specifically being aware of the mass of the universe. Evidently no one is actually aware of the mass of the universe, and, as Socrates surmised, perhaps no one is really wise. Accordingly, 'wise' and 'aware of the mass of the universe' are coextensive, nobody is one but not the other. The two differ, however, with respect to concomitance. There may be a certain wisdom that comes of being in awe of the mass of the universe, but minor increases in such awareness do not parallel differences in wisdom, in extents of being wise or unwise. The degree of sameness or difference is reflected by the degree of concomitance. Wisdom waxes and wanes, e.g., with experience or romantic insanity, but not in unison with the degree of one's awareness of what the mass of the universe is. Given merely that Meno was not as wise as Socrates and that both were equally unaware of the mass of the universe, it follows that the two properties are not concomitant, and hence neither the same nor synonymous, regardless of being coextensive.

Concomitant variation provides a much finer filter for testing meaning hypotheses than coextensiveness, a tool which is more and more powerful with closer and closer study of differences in extent. If coextensiveness is ultimately the only test imagined for determining translation, it is perhaps not surprising that translation should seem "indeterminate." Despite being coextensive, 'wise' and 'aware of the mass of the universe' differ radically in meaning, and, not by accident, diverge radically with respect to concomitance.

Philosophers have tended to overlook the potential of degrees and

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

extents to individuate properties or divine real differences in meaning. In semantic theory dominated by the concepts of reference and extension, there is little hint of the sensitivity of degrees to semantic difference or extensional anomaly.

Properties that are inherently connected may not diverge dramatically with respect to concomitance. Differences in expression, syntax, or spelling may not be detected in either extents or extensions. No difference between 'scepticism' and 'skepticism' can be expected. Grammar and syntax, like most things, admit of degree. Being decreased and being diminished are coextensive properties that may also be concomitant. Some differences in meaning may elude both tests. If 'being synonymous' and 'having the same meaning' themselves possess some subtle difference in meaning, it is not one clearly captured by any difference in extension or extent. As a test for synonymy, sameness of degree structure (coextensiveness and concomitance), though potent, may be imperfect. Translation, matching meanings, may not be utterly determinate, but it may not be as radically indeterminate as reflections on the limits of coextensiveness might lead us to suppose.

Concomitance provides an indirect test of modity. Given, for example, that being empirical means being, as Hilary Putnam[15] suggests, "subject to experimental or observational test" and that being susceptible to such test admits of degree, it follows that either the matter of being empirical is mod or else the analysis assumed is incorrect. Conversely, the modity of being empirical may be sufficiently clear to cast doubt on any analysis in which it is not explained as mod.

In general, if being A is analyzed as being B, then A and B should be concomitant. Coextensiveness may be difficult to judge without clear, unequivocal cases, e.g., definitely A but not–B, or definitely B but not–A. Definitive cases may be elusive if the matter is one subject to philosophical dispute. Concomitance, on the other hand, can be assessed without determining how the A/not–A dichotomy might be drawn.

If justice is fairness, beauty truth, or ignorance bliss, then degrees and extents must match. If justice really is fairness, then any adjustment in a social arrangement which modulates the one should identically modulate the other; any which diminishes the one should, to the same degree, abate the other. This is a severe test, but one easier to administer impartially than debating coextensiveness, with a polemical seesaw over any proposed classification: unfair or unjust. Uncontroversial instances are generally scarce, but differences in extent

[15] *Mathematics, Matter, and Method,* vol. I (New York: Cambridge, 1975), p. vii.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

are unlimited. If confirmation is confined to ideal cases, conviction is apt to be less than ideal.

As analytic tools, coextensiveness and the counterexample method have the drawback that, if we are not already conclusive about what is or is not *A*, we cannot then be conclusive about what is *A* but not–*B*. It seems we are trapped in the paradox of analysis. Concomitance, however, can be tested without preanalytically defined cases, without presuming that we already know what is or is not–*A*. Looking to coherence considerations to undo the paradox brings scant relief, if our surrounding theory embodies only low-grade generalizations concerning the simple presence or absence of all–or–nothing properties.

A theory or an analysis of *A*-ity, whatever else its merits, is incomplete even if it succeeds in resolving necessary and sufficient conditions for 'is *A*'; the result is incorrect if the degree structures do not match concomitantly, and incomplete if degrees and extents are undisclosed. The seemingly simple matter of introducing a novel term via an explicit definition may be incomplete without further explanation. For example, Nelson Goodman's[16] 'grue' is semantically defective in a way that 'blue' and 'green' are not. There are no specifications, established usages, or conventions for determining 'as grue' and 'not as grue'. Does becoming gruer mean becoming greener, bluer, softer, sweeter? It might be relevant to know how differences in temperature, time, color, pressure, etc., relate, whether, e.g., emeralds becoming bluer and bluer as we approach closer and closer to time *t* means they are becoming more and more grue or less and less grue. Would an emerald turning less and less grue be closer and closer to turning grue? How blue must an emerald after *t* become to be gruer than an emerald before *t*? We do not know, 'grue' as such is undefined. Perhaps an account of 'grue' can be given in which the orderings are provided, coherently disentangled, and 'grue' is no more semantically defective than 'green' or 'blue'. But, without such an account, novel terms such as 'grue' are inadequately or only partially defined.

VI. FREEDOM AND RESPONSIBILITY

Not attending to degrees, difficult problems may be more bewildering than necessary, like attempting to assemble a puzzle while ignoring the shapes of the pieces. For example, consider the problem of moral responsibility and freedom. According to a forceful moral intuition, a person cannot be morally responsible unless free to have done otherwise; the moral 'ought' implies 'can'. How then can de-

[16] *Fact, Fiction, and Forecast* (Indianapolis: Bobbs-Merrill, 1965), pp. 74–80. That 'grue' might have been more fully defined and Goodman's point remain is not at issue.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

terminism be true, a person not be perfectly free, and yet be morally responsible? How can these opposing considerations be reconciled? There is a natural answer given the modity of the matters involved. How responsible one can be held to be is limited by how free one was to have done otherwise. The basic principle is aptly understood as an extent principle specifying that

(FR) one is responsible only to the extent one chose freely, the extent one was free to have done otherwise.

Extent of responsibility is accordingly less where one is less free. Freedom is extremely mod, and there may not be any such condition as being perfectly free, no more than there is of being perfectly distant. If freedom does not admit of an intelligible perfect degree, the notion of perfectly or absolutely free cannot define the distinction between being free and being not free. There is not apt to be any sharp free/not–free division. A person free to choose according to contemplation, whim, or fancy is freer than if coerced, freer to the extent not intimidated, freer to the extent not disinformed, threatened, terrorized, or shoved. The relative differences do not depend on there being absolute instances of being utterly free. Possibly even our whims and fancies are traceable to the distant past, to peculiarities of nature or nurture, or to demon or divine creation. Perhaps no act is perfectly free, but again 'perfectly free' is not a notion that need be taken seriously. Confusion sets in if we imagine that freedom must be gauged by approximation to some absolute standard of being free to an utmost possible degree, a notion no more intelligible than that distance must be gauged by approximation to some absolute point perfectly distant. The situation deteriorates if being free is identified with being perfectly free.

Searching for a free/not–free dichotomy may be futile. The principle formulated concerning extents (FR) is independent of weary struggles over setting such a division. Degree of responsibility is adjusted in proportion to degree of freedom. Free/not–free questions may be too vague to admit of any reasonable resolution. Whichever way the free/not–free debate is decided, the pertinent questions of extent are unaffected. How free or unfree is not determined by locating a free/not–free dichotomy. For example, even if all instances of blushing, blinking, browsing, embezzling, and burglarizing are determined to be alike in being not–free, it does not follow that they are always free or unfree to the same extent.

Conceiving of 'free' and 'not–free' as labels for kinds masks important differences. There are apt to be serious differences in extent among actions deemed free, and among actions judged not–free. It is only such differences that are relevant to the principle in the

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

extent form. Applying a 'not–free' label, however final, does not eliminate these differences. That determinism is incompatible with being perfectly free should be no more incapacitating than that being somewhere is incompatible with being perfectly distant.

Initially dealing with degrees may appear more complicated, but ultimately there is little alternative. Dwelling on ideal cases, aside from questions of relevance, may be a misleading encumbrance. Attempting to treat concepts such as freedom and responsibility as nonmod, all-or-nothing properties, we encounter an unnecessary degree of theoretical fiction, one that makes complicated matters even more confounding.

It is not, as we have seen, essential for meaningfulness that there be a fine line deciding when something is or is not–*A*; and it is not a requirement of reason that such lines must exist or be legislated when none can be discerned. Properties change, the most abrupt changes may be somewhat gradual. The so–called "fine line" between this and that which we hear of so often may simply be a colloquial myth, it is not a conceptual necessity. It may be that personhood comes gradually, that conception itself is gradual, that there was no first human, no first giraffe, first toxic increment, biped, word, or moral agent.[17] The principles for such matters may be only those which relate differences in extent. Perhaps by nature most matters are continuous, sloped, and only gradually cleft. Assuming or idealizing otherwise, we may in the end manage only to misconceive them. The practice in general of treating matters of degree merely as matters of kind may be unsound, with no serious advantage in constructing principles and theories that do not employ or heed extents and degrees. Considering our initial hypothesis, it is not surprising that problems arise. In omitting degrees, it is inevitable that the resulting definitions, analyses, and theories are incomplete. Whatever explanatory power we might wish a theory to possess, it is limited to the extent that differences in degree are left mysterious. The practice of neglecting degrees and extents may be a tacit tradition, it is not a methodological imperative.

RAYME E. ENGEL

University of Houston

[17] Induction principles discredited by the apparent lack of such firsts contrast with degree principles that heed decreases in extent; an account is given in my "Properties, Degrees, and Induction." Such principles acknowledge, for example, that a property P over a series is unceasing if extents are preserved undiminished, if the $n + 1$th of the series is always as P as the nth in the series.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

COMMENTS AND CRITICISM

PERCEPTUAL PRESENTATIONS AND BIOLOGICAL FUNCTION: A COMMENT ON MATTHEN

IN "Biological Functions and Perceptual Content,"[1] Mohan Matthen claims that "any widely acceptable account of computational theories of perception must explicate what it means to endow a material state with representational content." And he attempts "to develop a theory of perceptual content which meets this requirement" (6). Matthen is correct to set this requirement on computational theories of perception. But, I shall contend, his own account fails to meet it.

To show this, it is convenient to begin with a misunderstanding Matthen has about my views. In "Intention and Action among the Macromolecules,"[2] I argued that the biochemist's use of intentional expressions to describe the behavior of macromolecules is so natural and unavoidable that it cannot be treated as mere metaphor. Matthen cites the passage, and attributes to me the conclusion that "the analogy between goal-directed behavior 'among the macromolecules' and perception, intention, and action in humans is enough to justify or even compel the use of intentionality" (18). Matthen goes on to say that he disagrees with this conclusion. So do I, for it was not my conclusion at all.

Having shown the naturalness of describing macromolecules as engaged in behavior like "recognizing, discriminating, undertaking tasks, making errors, editing, correcting, proofreading," I argued that such description could not be dispensed with in molecular biology, except at great inconvenience, too great for this vocabulary to be understood as convenient metaphor. I then noted, however, that attributions of these apparently intentional properties to macromolecules are thoroughly *extensional:* they remain true under all substitutions of coextensive predicates and coreferring terms in the biological descriptions where they figure. Accordingly, I concluded, these apparently intentional terms are implicitly redefined by biologists so as to be no longer intentional. Accordingly, the naturalness

[1] This JOURNAL, LXXXV, 1 (January 1988): 5–27. Unless otherwise noted, all page references are to this paper.

[2] In Nicolas Rescher, ed., *Current Issues in Teleology* (Lanham, MD: University Press of America, 1986), pp. 65–76. A revised version of this argument appears in my *The Structure of Biological Science* (New York: Cambridge, 1985), ch. 8.

0022-362X/89/8601/38–44
© 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

of intentional attributions in molecular biology does not "justify or even compel the use of intentionality." It avoids it altogether.

My argument exploited a biochemical description of the activities of transfer RNA. But this implicit redefinition can be illustrated in Matthen's own example of the "quasi-perceptual states" of a molecular "operon" (16–19). This is a portion of the genetic material which switches on and off the production of ribosomes, the organelles that synthesize proteins in the cell. Matthen accords operons quasi-perceptual states, because they "sense" when ribosome synthesis is going too fast or too slow, and act to decrease or increase this synthesis. But the operon "senses" these processes under any substitution of coextensive or coreferring terms in the description of ribosome synthesis, and, furthermore, there is an available, purely extensional description of the process we are inclined to describe as sensing the presence of excess or insufficient ribosomes. So, there is no intenSionality[3] here, and consequently no intentionality.

It is an obvious premise of this argument that intentionality requires intenSionality. This premise is widely shared in the philosophy of psychology.[4] Indeed, since Franz Brentano it has been argued that intenSionality is a mark of the mental. There are several arguments for this view. I believe the most powerful one is this: intentional states are intenSional precisely because they are *representational,* because they bring their objects under descriptions, and thus involve language, whether a natural one, or a language of thought.

But it is just this feature of intentionality which Matthen's account of representational content in perception fails to accommodate. And it fails to do so because of features it shares with other "naturalistic" theories of psychological content, like those of Daniel Dennett[5] and Jonathan Bennett.[6] Precisely like these older approaches, Matthen's theory purports to analyze psychological content into a species of fine-grained teleology. In Dennett's approach, neural networks are accorded "content" in virtue of the environmental appropriateness of their afferents to their efferents; and environmental appropriateness is explicated in terms of blind variation and natural selection. In Bennett's theory, organisms are said to "register" statements about their environments just to the extent such registrations are required

[3] I capitalize the 'S' to emphasize the difference. I owe this orthographic innovation to Kathleen Emmett.

[4] This is a premise Matthen shares. He rejects causal accounts of the "sensing" of operons, because they do "not . . . give us a schema that is potentially able to handle the intensionality of perception" (19). Neither does his functional account, however, as I shall argue.

[5] *Content and Consciousness* (London: Routledge & Kegan Paul, 1969).

[6] *Linguistic Behavior* (New York: Cambridge, 1976).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

to explain their behavior in the light of their goals, where the goals are given an evolutionary explanation. Both Dennett and Bennett acknowledge that the sort of content attributable to organisms under this strategy is insufficient to identify any particular statement as *the* content of a neural process or a registration, and Bennett has even admitted that, on this approach, registrations will not be sensitive to substitutions of truth-functionally equivalent statements in the description of their contents (*op. cit.*, pp. 113–116). This is one reason why Dennett eventually came to adopt "the Intentional Stance,"[7] according to which the attribution of intentional states is not be taken literally, but is a heuristic device.[8]

Matthen recapitulates the analytical strategies of Dennett and Bennett, but without making any further accommodation for intentionality than theirs do. Thus, it is ultimately no more satisfactory.

Matthen clearly recognizes that perceptual states will have to be representational and intentional: "the content of a perceptual state is . . . *expressible* by means of a definite description. Then the object of perception is the unique object that satisfies the definite description . . . the content of a perceptual state . . . *presents* an object as possessing a certain characteristic" (7, emphasis in original). And Matthen notes that the pressing question for the philosophy of psychology is: "What feature of a perceptual state makes it a presentation of a particular property, say, redness?" (7).

To answer this question, Matthen employs Wright's analysis of functions.[9] Matthen defines a "quasi-perceptual state," and then employs this notion to define a full perceptual state. But neither reflects the appropriate intenSionality. A quasi-perceptual state, P, is one which brings about an appropriate response. P is, in Matthen's terms, a "presentation" of a certain property, F, of the items it responds to, if and only if some characteristic of P "bring[s] about the functionally appropriate response to the presence of things that are F." But, on Wright's account, or any other account of function for that matter, whatever brings about an appropriate response to things with the property Fx, will bring it about to things with the property Fx v Gx, or any other co-extensive property.

[7] "Intentional Systems," this JOURNAL, LXVIII, 4 (February 25, 1971): 87–106; reprinted in *Brainstorms* (Montgomery, VT: Bradford, 1979).

[8] For a detailed examination of the failure of naturalistic strategies to explicate intenSionality, see my "Evolutionary Biology and Intentional Psychology—Part I: The Uneasy Analogy," "Part II: the Crucial Disanalogy," *Behaviorism*, XIV (1987): 15–28, 125–138.

[9] Larry Wright, "Functions," *Philosophical Review*, LXXXII, 2 (1973): 139–148.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

On Wright's analysis, as adapted by Matten, *G* has the function *F* iff

(1) *G* has the consequence *F* in circumstances *C*.
(2) *G* is there because it does *F* in circumstances *C*.

The second clause obtains through the operation of selection over variation in the evolutionary past. But, if *G* is there because it has been selected for doing *F*, it is there because it has been naturally selected for doing *F* under any extensionally equivalent description of doing *F*.

Nothing reflects this fact about functional statements more clearly than examples from molecular biology, like my DNA proofreading enzyme example or Matthen's Ribosome-sensing operon. In these cases, we know several alternative descriptions of the effect a functional system has, we have a complete causal account of how it accomplishes this effect, and we can identify the system independently of its effects. Here everything is transparent, and more important, known to be transparent. So there is no scope for intenSionality in the attribution of function.

Thus, there is no single property that is unambiguously or definitely "presented" in the functionally appropriate response of some functional system, or even a finite number of them. Therefore, no particular content can be unambiguously attributed to what Matthen defines as a quasi-perceptual state. There is, in short, no particular feature of the environment which is represented in the quasi-perceptual state *S*.

Matthen's definition of a full perceptual state does not make good this omission. In fact, it may repeat it. He notes that the definition of quasi-perceptual state "has one very obvious disadvantage when . . . applied to real perception such as we humans experience." It is that there is no unique functionally appropriate response among humans to "what a perception tells us." But this unique response is built into the definition of quasi perception. So, Matthen's definition of the sort of perception humans engage in substitutes for "bringing about a functionally appropriate response" the function of "*detecting* the presence of things of a certain type" (20, emphasis in original). But what is "*detecting* the presence of a thing with property *F*"? If detecting is an intentional state, then the definition of perception is unhelpful, indeed it fails the demand that Matthen lays down: that an analysis "must explicate what it means to endow a material state with representational content," for it merely substitutes one intentional state for another in the worst traditions of homuncularism. And, if

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

detection is not a matter of internal representation, but of appropriate behavior, the definition does not avoid the "disadvantage" Matthen himself identifies for his account of quasi perception: among humans there is no such thing as *the* behavior appropriate to "what the perception tells us."

Matthen claims that a "strength of the functional schema [he provides for perception] is that it accounts well for the peculiarities of intentional explanation—the explanation of the causal interrelations of states by reference to the content assigned to them." This would be an important strength, if Matthen can show how functional states can have content—psychological content. In a sense, functional states have "content": they have adaptational *significance* or *meaning,* but this kind of "content" is purely extensional, and so is not relevant to the kind of psychological content in question here.

The "content" of a functional state is given by those of its effects which are adaptive, or have been in the evolutionary past. Like psychological states, adaptational "content" is supervenient on structural properties. Thus, two different organisms can have the same level of evolutionary fitness, even though they are structurally dissimilar and even differ in ambient environments. For example, wings and legs might both have the same adaptive significance because of their functionally identical role in predator escape. But this kind of evolutionary content is not a matter of what Matthen calls "abstract semantic relations in the realm of propositions." As the preservation of truth under substitution of coextensive terms shows in these cases, the functional states of biology are connected to their causes and effects in ways that do not generate intenSionality, and so will not in the end account for intentionality.

Intentionality cannot simply be a matter of functional organization, because such organization never requires a representation of the effect which the organization attains or at which it aims. But, perhaps intentionality is causally or conceptually prior to functional organization, and somehow transmitted through it; functionally organized systems betray "derived" intentionality, because their components or whatever put them together bear it. This at any rate seems to be Dennett's latest view. Perhaps it can help Matthen's project. After quoting the same passages from "Intention and Action among the Macromolecules" which Matthen treats, Dennett[10] writes:

> The disappearance of intenSionality at the macromolecular level at first seems a telling objection to the persistent use of intentional idioms to

[10] *The Intentional Stance* (Cambridge: MIT, 1987).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

characterize that level, but if we leave it at that we miss a still deeper level at which the missing intenSionality reappears (*ibid.*, p. 317).

Dennett's argument is roughly this. Enzymes do not have reasons for what they do. But nature does. It is *natural selection* that recognizes a reason for the enzymes and operons to do what they do, and that designs them accordingly:

> If we work out the rationales of these bits of organic genius, we will be left having to attribute—but not in any mysterious way—an emergent appreciation or recognition of those rationales to mother nature herself. . . . We may call our own intentionality real, but we must recognize that it is derived from the intentionality of natural selection, which is just as real—but just less easily discerned because of the vast difference in time scale and size (*ibid.*, pp. 317/8).

Now, there are two ways to treat this claim of Dennett's: either it is metaphor or it is a literal claim that there are free floating rationales or intentions independent of sentient creatures to which we accord psychological states. If it is metaphor, then obviously it cannot do Matthen's theory much good. For then the most it can show is that the intentionality of psychological states is metaphorical—the reflection of a stance we adopt to predict the behavior of complex organisms and to guide the construction of nonintensional theories of their behavior. And, if it is literal, then we have not, after all, met the demand with which Matten begins his project: "explicating what it means to endow a material state with representational content" (6).

When Dennett says that the intentionality of natural selection is real, I think he mispeaks himself. For he goes on in the next paragraph to deprive nature of the very thing that real intentionality requires:

> . . . if there is to be any original intentionality—original just in the sense of being derived from no other, ulterior source—the intentionality of natural selection deserves the honor. What is particularly satisfying about this is that we end the threatened regress of derivation with something of the right metaphysical sort: a *blind* and *unrepresenting* source of our own sightful and insightful powers of representation (*ibid.*, p. 318).

I take this to be the admission that intentionality of all sorts is metaphorical,[11] a stance we adopt for our own convenience, and so not

[11] Dennett denies it is metaphorical, though he insists that it is a stance. See *The Intentional Stance*, pp. 55–57. Here Dennett treats intentional states as what he calls "abstracta," concepts like "center of gravity," which are "objective" heuristic devices.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

something of which we need to explain the actuality. Treating intentionality as a heuristic device is the inevitable conclusion of any attempt to derive real intentionality from functional organization. The most we can expect a purely functionalist account of perception to show is how a material state can mimic, or asymptotically approach, representational content. If so, there is no patching up that will overcome the problems besetting Matthen's theory of perceptual content.

ALEXANDER ROSENBERG

University of California/Riverside

BOOK REVIEWS

Descartes. JOHN COTTINGHAM. New York: Basil Blackwell, 1986. 171 p. Cloth $45.00, paper $14.95.

Cottingham's *Descartes* has the "modest" aims of giving "a clear and straightforward account of Descartes's philosophy which does justice to the complexities of argument involved, while at the same time attempting to make the issues reasonably accessible to students who may be approaching the Cartesian system for the first time" (vii). Cottingham is not intimidated by the difficulty of the task; accessibility need not, he says, mean oversimplification (vii). Descartes did hold that no aspect of his philosophy should be unintelligible to an untrained mind approaching it without preconceived opinions, provided only the reasoning was laid out simply enough and in the appropriate order. But could such a brief account of Descartes's philosophy be both clear and accessible and still do justice to the complexity of the issues and arguments involved? For a reader having any acquaintance with Descartes's own writings, it soon becomes apparent that the ideal of accessibility Cottingham aims at cannot be attained without a certain measure of radical oversimplification.

Clarity is, without doubt, one of the virtues of Cottingham's account, and it provides a useful and honest basic introduction to the best-known themes of Descartes's philosophy. It is most profitable when read with the texts, either the Adam & Tannery edition, or, for those not having access to it, the new Cottingham, Stoothoff, and Murdoch edition to which it obviously constitutes a companion.[1] The

[1] *Oeuvres de Descartes,* C. Adam and P. Tannery, eds. (Paris: Cerf, 1897–1913) (repr. Paris: Vrin, 1964–1976), 12 vols. (quoted as AT); and *The Philosophical Works of Descartes,* J. Cottingham, R. Stoothoff, and D. Murdoch, trans. (New York: Cambridge, 1985), 2 vols. (quoted as CSM).

0022-362X/89/8601/44–48 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

references to contemporary debates on standard controversial questions contribute to its value as a guide to recent discussions, restricted, however, mainly to Anglo-American scholarship. Great weight is given, on controversial issues, to the testimony of Burman. This is natural given the general aim of the book, and Descartes's English-speaking readers have Cottingham to thank for a valuable edition of the *Conversation with Burman.*[2] It is also symptomatic; the very character of an informal (dinner?) conversation reported retrospectively suits well the standard of accessibility set by Cottingham.

Cottingham avoids detailed questions of exegesis and sophistication. There is an undeniable, innocent-looking charm in his own way of reading Descartes, which could be described as an application of Ockham's razor to the multiplicity of interpretations: never complicate the thought of an author without necessity. If it is possible to find a coherent reading on disputed topics which does not stray too far from what passes as commonsense in present–day scientific communities, or—when, as in metaphysics, commonsense is not applicable—from received opinions on Descartes's philosophy and its incoherencies, we can be sure to find it in Cottingham. Although this way of proceeding may fit well his general purpose, when the issues are complex, it runs the risk of leading seriously astray. Descartes, after all, is anything but a commonsense philosopher, and, in spite of his repeated confessions that truth is simple and evident and his philosophy merely an effort to remind us of what has always been known, what is simple in philosophy can be very hard to see, and this is something of which Cottingham sometimes tends to loose sight.

Putting it crudely, the Descartes emerging from this book is more a scientific reformer with philosophical afterthoughts due to obsessive doubts about the reality of the objects of his investigations, than a great philosopher confronting the epistemological and conceptual challenges raised by the new science. The famous "method of doubt," for instance, summarized first in the *Discourse* and fully developed in the *Meditations,* is described as the outcome of a "new and much more dynamic philosophical approach" compared to "the unsatisfyingly static quality" of the methodological doctrine outlined in the earlier *Rules* (28). Although Cottingham does give a brief account of intuition and stresses its central role in the Cartesian quest for true knowledge, nothing is said about the deep and far-reaching difficulties involved in Descartes's methodological ideal. The method of doubt is presented as just one development of "germs" already present in the *Rules,* and the need to provide a

[2] New York: Oxford, 1976.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

secure foundation for the new science of nature is seen as an accidental psychological feature (e.g., excessive ambition [73]) of one of its proponents, rather than an issue inherent in Descartes's very project of a unified science based on direct mental intuition of its objects. Without further explanation, all this Cartesian fuss about proving the existence of the self and the external world, as well as the worries about warranting self-evident truths by divine authority, certainly must look extravagant to present-day students, as it did to Descartes's Aristotelian and empiricist contemporaries. More attention could have been given to what Descartes states as his general aim throughout the *Meditations:* that of leading the mind away from the senses and proving the superiority of the intellect over the senses as the only source of certain knowledge. (Synopsis, AT VII, 12–16; CSM II, 9–11; Cottingham, 31 ff.).

Because he does not give enough weight to the distinction drawn by Descartes between the senses and the intellect, Cottingham fails to grasp the role and import of the malicious demon argument. That this argument may have some other purpose than reinforcing doubt about the reality of the external world is hinted at in a note (33, fn. 23) but not seriously considered. Cottingham rejects, as an unfounded "supposition," the possibility of Cartesian doubt reaching out to reason itself and to the very foundation of rational principles on which the knowledge of reality is based, and hence misses one of the central and philosophically most profound implications of Descartes's methodological skepticism. Cartesian doubt is seen as a mere vehicle or "mechanism" for providing secure first principles (28, 42). Cottingham is right, of course, that Descartes's business is not to contribute to the philosophical debate over skepticism, and also in insisting on the temporary, methodological and hyperbolical character of the skeptical doubts used by Descartes (43). He refers to Descartes's own testimony that the value of the counterarguments given against them is not "that they prove what they establish—namely, that there really is a world, and that human beings have bodies and so on . . . ," things that "no sane person has ever seriously doubted" (Synopsis, AT VII, 15/6; CSM II, 11; quoted by Cottingham, p. 34). But he overlooks the very point of the test to which the arguments proving the existence of the (material) world and bodies are put, which is to show that they are less "solid and transparent" than those leading to the knowledge of our own minds and of God. He does not, therefore, take account sufficiently of what Descartes in the very same text declares to be the sole aim of the *Meditations:* to establish that the latter, which can be known only through the intellect, are the most certain and evident of all possible objects of human knowledge (AT VII, 16; CSM II, 11).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Now, if the intellect is preferred to the senses as the only reliable source of knowledge, it is because the principles discovered through reason provide a more stable foundation for knowledge than the ever changing evidence of the senses. But Descartes is not just another Platonist. In fact, his doctrine of the creation of the eternal truths, which is another central tenet of Descartes's seriously underrated by Cottingham, renders his mathematical philosophy of nature rather different from that of the Platonists. The very reliability or permanence of the principles of reason cannot be taken for granted, but has to be warranted. This is what Descartes tries to do, by using the same method of doubt through which sense-based knowledge is tested. At this point, the traditional skeptical doubts have to be supplemented by the extravagant supposition of the malicious demon. Even in the famous cogito argument, Descartes's aim is not merely, as Cottingham claims, to establish the irrefutable certainty of his own existence (cf. 46, fn. 26), but to show that this certainty is acquired through the intellect alone. The logic of his own method then brings Descartes to question the very principle or truth of reason instantiated in the reasoning establishing his own existence, and, more generally, the validity of any rational principle. Cottingham does not see that there is any problem for Descartes about the foundation of truths of reason or mathematics. Descartes saw no other ground for their necessity than the unchanging divine will, hence the urgency to prove God's existence, and hence also all this trouble with the famous circle from which it is so very hard to see how Descartes really can escape. For Cottingham, there is no real circle, since the truths of reason are never questioned, and the whole business of God guaranteeing the rule of evidence is merely rooted in a worry about the limited scope of our human capacity of attention (70–72).

Cottingham fulfills his aims best in chapter 3 ("From Self to God to Knowledge of the World") when discussing Descartes's proofs of God's existence (although his criticism of what he calls the *causal adequacy principle* is based on a rendering of Descartes's principle of causality which is not adequate; 49 ff.), and in chapter 4. ("The Material Universe") where he accounts for Descartes's view of matter and physical science. The most interesting and also controversial chapters are the last two ones dealing with "The Cartesian Man" and "The Human Condition." The former includes material from an earlier paper, "Cartesian Trialism."[3] Departing from standard presentations of Cartesian dualism, Cottingham discusses the problems relating to Descartes's account of the body-dependent mental facul-

[3] *Mind*, XCIV, 374 (April 1985):218–230.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ties, imagination and sensation, and the implications his recognition of such complex psycho-physical phenomena have for his view of our knowledge of human nature. Their peculiar and inherently confused, phenomenological character not only proves, as Cottingham rightly stresses, the embodiment of the human mind, but also marks the limit of what can be conceived in terms of distinct concepts of thought and extension. The "trialism" emerging from Descartes's account of man as a union of mind and body which can be understood only in terms of a third primitive notion certainly conflicts with the standard reading of Cartesian dualism as a two–substance view, celebrated by Gilbert Ryle as the "official" doctrine of the mental and the physical instantiated in the allegedly Cartesian "Ghost-in-the-Machine" myth. But does Cottingham not accord too much to this reading in insisting on a distinction between what he calls Descartes's "official" dualist view and the "trialism" defended in his unpublished writings (130 ff.)? It is not difficult to find evidence in all of Descartes's major published works that he consistently held man to be not an accidental composite but a real "substantial" union of mind and body, described in traditional Aristotelian terms as a union of form and matter. That the mind is embodied is a given, initial fact for Descartes, which is commonly experienced and so obvious that it requires no argument. What needs to be proven, from Descartes's point of view (as so many of Descartes's French commentators have shown), is not the union but the real distinction between mind and body.

This is not to deny that Descartes's third primitive notion is problematic, given his radical distinction between the notions of mind and body, and Cottingham is right in stressing the "deep tension between what our reason tells us and what our ordinary everyday experience tells us" (132). Because of the nontransparent and ultimately unintelligible character of the thoughts depending on the mind-body union, which are "absolutely fundamental to what it is to be a human being," "a major part of our human life remains, in the end, a mystery" (132). Whether the fact that so much of human experience is left outside the scope of rational, scientific explanation renders Descartes's account of human nature unsuccessful, as Cottingham thinks (131), depends in part on what credit and value one is willing to accord scientific explanations in this domain. Those who are not happy with Hobbes's crude materialism surely rejoice that Descartes did not believe human thinking and experience to be explainable mechanistically, in terms of the shape, position, and motion of particles in the brain. Cottingham points out elsewhere how Descartes, in spite of his confidence in our divinely implanted mental capacities, repeatedly acknowledged the imperfect, limited character of human

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

knowledge (98). Why not credit Descartes with the insight that not only the ends of God in creating the universe, but also those dictating the laws instituting human nature, remain partly beyond our grasp?

It is not possible to go into all the interesting and difficult questions raised by Cottingham's discussion of Descartes's view of man and human cognition. His account of sensations, sense perception, and innate ideas constitutes a valuable contribution to discussions of Descartes's theory of knowledge, even if Cottingham in criticizing Descartes seems overly optimistic about what has or can be established by scientific research. The last two brief sections, "Freedom and Reason" and "The Good Life," complete the picture of the human condition. Cottingham obviously approves Descartes's prudential rationalism. But his account does not, in my view, do justice to Descartes's radical theory of freedom of the will. Cottingham does not show why the interpretation of Descartes as holding the same kind of view Spinoza and Leibniz held is "the most plausible interpretation" of his controversial statements about a radical freedom of choice (149–151, fn. 30). Neither Spinoza nor Leibniz found themselves, and for good reasons, in agreement with Descartes on this point, and in so far as the textual evidence goes, it can be argued that any reading attributing to Descartes some version of the traditional ethical determinism defended by other rationalists is rather implausible.

For those interested in the psychology of René Descartes, it should be noted that Cottingham includes, in an appendix to the book, a translation of Descartes's dramatic dreams of 1619 as reported in the biography of Adrien Baillet (1691).

LILLI ALANEN

University of Helsinki

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

NOTES AND NEWS

The editors wish to remind our correspondents that notices of scheduled meetings and symposia are of value to our readers only if they appear early enough to make attendance possible. Copy for *Notes and News* must be planned and processed in this office in time for it to go to the printer at least six weeks before publication date, the 15th of each month. Accordingly, an announcement of a meeting to be held, say, in April, should be scheduled for our February issue (or sooner) and should be received in this office no later than December 1, that is, at least four months in advance. The editors appreciate notices that come to us ready for the printer, requiring little editorial work and no retyping, in JOURNAL style (see these columns), with no imperatives, no prices, no telephone numbers, no headings, no unnecessary capitals; and reasonably terse, since space is limited.

The Hume Society will hold its Seventeenth Hume Conference at the Australian National University in Canberra, June 27–July 1, 1990. Papers may be submitted no later than September 1. Proposals, which should include the name of the proposer and a short list of the proposer's recent work relating to Hume or his historical context, will be considered in lieu of papers. Although the Program Committee will give full consideration to papers or proposals on any appropriate topic, it hopes to include symposia or continuing discussions of three themes, and wishes to encourage the submission of papers that contribute to the following themes: Hume and the Controversy over Infinite Divisibility; Personal Identity; Passions and Virtues; and Hume and the Foundation of Morals and Politics. Triplicate copies of proposals or papers may be directed to Dorthy Coleman, Phil. Dept., Bowdoin College, Brunswick, ME 04011.

In celebration of its 375th anniversary, Gronigen University will sponsor a conference on Knowledge and Language to be held May 22–24. Speakers will include Noam Chomsky (MIT), Arthur C. Danto (Columbia), Edward Said (Columbia), Mary Hesse (Cambridge), Harry Bracken (McGill), Kuno Lorenz (Saarbrucken), Ken Wexler (Irvine), Willem Levelt (Max Planck), Barbara Partee (Amherst), Dan Sperber (Univ. Paris X), Ray Jackendoff (Brandeis), Hayden White (UCSC), David Pesetsky (MIT), Richard Wollheim (London), and Michael Jorden (MIT). Further information may be obtained by writing Werner Abraham, Letteren, Rijksuniversiteit Gronigen, Gr. Rozenstraat 15, NL-9765 TG Gronigen, Holland.

The Norwegian University Press recently published an excellent paperback edition of Corliss Lamont's *The Philosophy of Humanism,* entitled *Det humanistiske livssyn.* This event shows the strength of humanist philosophy in Norway, which has the largest Humanist Society of any country in the world, with approximately 30,000 members.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The Eighteenth World Congress of Philosophy took place from August 21–27, 1988, in Brighton, England, organized by the Federation Internationale des Societes de Philosophie, with over one thousand delegates in attendance, from every region of the globe. The General Assembly of FISP voted to schedule the next World Congress for Moscow in 1993, and to hold an intermediate congress in Nairobi in 1991. In addition, the General Assembly formally encouraged the USA to continue planning for a 1998 World Congress in Atlanta. Five APA members were elected to serve ten–year terms on FISP's Steering Committee (joining Richard DeGeorge and Ruth Marcus, who are already members). Those newly elected are Anna Tymieniecka, Dagfinn Follesdal, Nicholas Rescher, Patrick Suppes, and Ernest Sosa (also elected to a five-year term as Vice President).

The Central States Philosophical Association will hold its 1989 meetings on October 13–14 at the University of Iowa. Philosophers from any geographical region are invited to submit papers by July 1. Inquiries may be directed to Panayot Butcharov, Phil. Dept., Univ. of Iowa, Iowa City, IA 52242.

The Conference for the Study of Political Thought will hold its 1989 International Meeting at Williams College, Williamstown, MA on April 28–29. The theme is "Markets and Political Theory." Major speakers will include Elizabeth Anderson, G. A. Cohen, Jean Hampton, David Miller, John Roemer, Alan Ryan, and others. Further information may be obtained by writing Charles Beitz, Political Science Dept., Swarthmore College, Swarthmore, PA 19081.

The General Editor of the critical edition of *The Works of George Santayana* would appreciate hearing from anyone with relevant material, such as letters, personal recollections, photographs, and unpublished writings.

Studies in Early Modern Philosophy, a journal concerned with the period in philosophy from the time of Francis Bacon (1561–1626) to that of Immanuel Kant (1724–1804), announces a call for papers. Papers may be directed to the Philosophy Dept., York Univ., Downsview, Ontario, Canada M3J 1P3.

On the occasion of the fifth anniversary of the death of Alfred Tarski, the University of California/Berkeley announces the inauguration of an annual series of Alfred Tarski Lectures. Dana S. Scott will begin this program by presenting three lectures during the week of March 27 on ideas that have evolved from Tarski's work in some of the many fields to which he contributed. Further information may be obtained by writing to the Tarski Committee, Mathematics Dept., Univ. of California, Berkeley, CA 94720.

The Department of Philosophy of the University of California/Irvine is cosponsoring a conference entitled "Life at the Limits: A Conference on the Moral Issues" to be held on February 24–25. The speakers and panelists will include Margaret Battin, Dan Brock, Allen Buchanan, David Daube, Philippa

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Foot, Louis Gluck, Sanford Kadish, Elisabeth Lloyd, Bernard Lo, Joanne Lynn, Deborah Mathieu, Warren Quinn, Leslie Rothenberg, T. M. Scanlon, Richard Wasserstrom, Bernard Williams, and Richard Wollheim. Further information may be obtained by writing to the Department, Univ. of California/Irvine, CA 92717.

The Department of Philosophy of the University of Maryland/College Park is pleased to announce the creation of a specialization in cognitive science, a consequence of the recent appointments of Michael Devitt as Professor, Christopher Cherniak and Georges Rey as Associate Professors, John Horty and Kenneth Taylor as Assistant Professors, and Norbert Hornstein as Affiliate Associate Professor.

The SUNY/Buffalo Philosophy Department will host a conference on the topic of "Law and the Legitimation of Violence," March 16–18. Speakers will include Hugo Bedau (Tufts), Sergio Cotta (Rome), John Ladd (Brown), Andre Maury (Helsinki), Jan Narveson (Waterloo), J. C. Nyiri (Budapest), Eike von Savigny (Bielefeld), Bernhard Waldenfels (Ruhr), Carl Wellman (St. Louis), and Elizabeth Wolgast (UC/Hayward). Papers and requests for additional information may be submitted to Newton Garver, Phil. Dept., Baldy Hall, SUNY/Buffalo, NY 14260.

The Department of Philosophy at Virginia Polytechnic Institute and State University announces a conference on "Liberty, Equality, and Fraternity" to be held April 21–23, on the occasion of the 200th anniversary of the French Revolution. Speakers will include T. E. Hill, Jr., Bruce Landesman, William Nelson, Kai Nielsen, and Gerald Postema. Further information may be obtained by contacting Harlan B. Miller, Phil. Dept., VPI & SU, Blacksburg, VA 24061.

The Department of Philosophy at Wake Forest University will hold its eighth James Montgomery Hester seminar on April 14–15. The topic will be "Skepticism, Faith, and Rationality." Speakers will include William Alston, Robert Audi, Terence Penelhum, and Richard Popkin. Further information may be obtained by contacting Gregory Pritchard, Phil. Dept., Box 7332, Wake Forest Univ., Winston-Salem, NC 27109.

The Department of Linguistics and Philosophy of the Massachusetts Institute of Technology is pleased to announce the appointments of Robert Stalnaker as Professor, and David O. Brink, Michael O. Hardimon, and Paul Hoffman as Assistant Professors.

The Department of Philosophy of Rice University is pleased to announce the appointments of Donald Morrison as Assistant Professor, and Martin Rice and Stephen Sullivan as Visiting Assistant Professors. In addition, Mark Kulstead has been appointed to succeed Richard E. Grandy as Chairman.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The Department of Philosophy of Rutgers University is pleased to announce the appointments of Professors Jerry Fodor and Barry Loewer as permanent members of the faculty.

The Department of Philosophy at the University of California/Davis, is pleased to announce the appointments of Michael Jubien as Professor, and Jean Hampton and Richard Healey as Associate Professors.

The Department of Philosophy at the University of California/Irvine is pleased to announce the appointment of Alan Nelson as Assistant Professor. The Department also wishes to call attention to its Distinguished Visiting Professorship, by which philosophers from different fields and various countries are brought to campus for a quarter or more each year. The Distinguished Visiting Professors for 1989 will be Peter Simons and Robert Sleigh.

The Department of Philosophy at the University of Colorado/Boulder is pleased to announce the appointments of Professor Dale Jamieson as Director of the Center for Values and Social Policy, Christopher Shields as Assistant Professor beginning in Fall 1988, and George Bealer as Professor beginning in Spring 1989.

The Department of Philosophy of the University of Kansas is pleased to announce the appointment of Ann Cudd as Assistant Professor, and James Muyskens as Professor and as Dean of Liberal Arts and Sciences.

The Department of Philosophy at the University of Minnesota/Duluth is proud to announce the appointments of Eve Browning Cole as Assistant Professor of Philosophy and Women's Studies, and Gerald Gaus as Associate Professor of Political Science and Philosophy.

The Department of Philosophy of the University of North Carolina/Greensboro is pleased to announce the appointment of Michael Zimmerman as Assistant Professor.

The Department of Philosophy of the University of Rochester is pleased to announce the appointments of Dorit Bar-on and Michele Moody-Adams as Assistant Professors, Randall Curren as Assistant Professor of Philosophy and Education, and Jeffrey Spike as Assistant Professor of Medical Humanities.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

MEMO TO AUTHORS

The editors cannot afford to send to the printer manuscripts that are difficult for the compositor to read or footnotes that need to be completed in proof. They ask that manuscripts submitted be prepared according to the following guidelines:

1. All material, including the quotations and footnotes—especially footnotes—must be typed, double-spaced, on one side of the paper, with ample margins left and right, top and bottom.
2. No bibliographies will be published; bibliographical material should be put into the footnotes.
3. Footnotes should be complete, according to *Journal* style (see any issue). Journal citations should include name of journal in full, volume number, issue number, month or season, year, and pages. Book references should include publisher, city of publication, and date. Footnotes should be gathered at the end of the article.
4. Quotations should be checked by the author; capitalization, italics, spelling, etc., should match the original. Corrections should not have to be made (expensively) in proof.
5. If there are diagrams, they should be provided camera-ready by the author. If there are special symbols, they should be widely spaced and clearly marked; if there are many, a list should be provided. Standard logic symbols are available, and letters of the greek and script alphabets; no boldface.
6. Computer copy can be extremely difficult to read; only letter-quality copy is acceptable. It should be double-spaced and not justified. The print should be dark enough and the letters widely enough spaced so that both letters and punctuation marks can be read easily.
7. Finally, photo copies on shiny paper are not acceptable. Ribbon copy is much preferred, but clear copies on bond paper are acceptable if necessary.

The editors are willing to read and evaluate manuscripts that do not meet these standards, provided they are genuinely legible and not unreasonably long (over 15000 words). They ask for revision of excessively long manuscripts, and they require retyping of accepted manuscripts, if necessary, before scheduling for publication.

The *Journal of Philosophy* is published monthly, in issues that (ordinarily) contain only 56 pages of copy; this limits the length of papers the editors can normally consider, to no more than 7,500 words, or eighteen JOURNAL pages. Manuscripts very much in excess of these limits will be returned unread.

The *Journal of Philosophy* does not publish papers that are primarily historical or expository; the editors urge that such manuscripts be submitted elsewhere.

CC-0. Mumukshu Bhawan Varanasi Collection. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

JOHNSONIAN PRIZE IN PHILOSOPHY

The editors are pleased to announce that the winner of the Johnsonian Prize in Philosophy for 1988 is Vann McGee, Assistant Professor of Philosophy, University of Arizona, for his manuscript, *Truth, Vagueness, and Paradox: An Essay on the Logic of Truth,* which will be published by Hackett Publishing, the co-sponsor of the prize.

Previous Johnsonian prize winners have been: 1980, Daniel Bonevac, *Reduction in the Abstract Sciences* (1981); 1981, Igal Kvart, *A Theory of Counterfactuals* (1986); 1982, Moreland Perkins, *Sensing the World* (1983); 1984, Robert J. Dunn, *The Possibility of Weakness of Will* (1986); 1985, Harry van der Linden, *Kantian Ethics and Socialism* (1988); and 1986, Clyde L. Hardin, *Color for Philosophers* (1988). Also, honorable mention in 1981 went to Charles B. Guignon, *Heidegger and the Problem of Knowledge* (1983); and in 1986 to Dionysios Anapolitanos, *Leibniz: Representation, Continuity, and the Spatiotemporal.*

The Johnsonian Prize, which consists of one thousand dollars and guaranteed publication of the prize-winning book, will be offered again in 1989. The prize is designed to encourage the work of unpublished philosophers; authors are eligible who have published no book-length work. Books are selected from among manuscripts sponsored by departments of philosophy. Authors and departments who wish to submit manuscripts should observe the following guidelines:

1. Any graduate school of philosophy that confers the Ph.D. degree may submit one manuscript per year. It is not required that the work sponsored by a department be written by either a graduate or a faculty member of that department; it may, but need not be a revised version of a departmental dissertation. The closing date for this year's prize is August 1, 1989.

2. To qualify, the manuscript must be: (a) of book length (at least 40,000 words) and in proper form for publication as a book; (b) accompanied by an abstract of up to 2000 words; (c) written by someone who has not yet published any book-length work; (d) accompanied by two professional evaluations and a letter from the sponsoring department (the editors are depending on the departments to screen out all but the most important and original work); (e) uncommitted to a publisher. Manuscripts will be returned to the authors only if accompanied by appropriate postage.

3. Graduate schools who wish to participate should notify the JOURNAL by May 15, 1989, of their intention to submit a manuscript for consideration, and, if possible, send evaluations and abstracts in advance.

4. Preference will be given to the work of younger philosophers, especially those who have not yet been granted tenure.

Announcement of the eighth award will be made in the January 1990 issue of this JOURNAL; some manuscripts may be given honorable mention. Publication of the prize-winning book will be guaranteed by the Hackett Publishing Company after the author has made what revisions are deemed necessary by the editors. The first six prize-winning books will be on display at the APA Eastern Division meetings in 1989.

PLEASE NOTE: The deadline for submission of manuscripts for the Johnsonian Prize for *next* year will be *January 15, 1989.* Announcement of the ninth award will appear in a Fall 1990 issue of this JOURNAL.

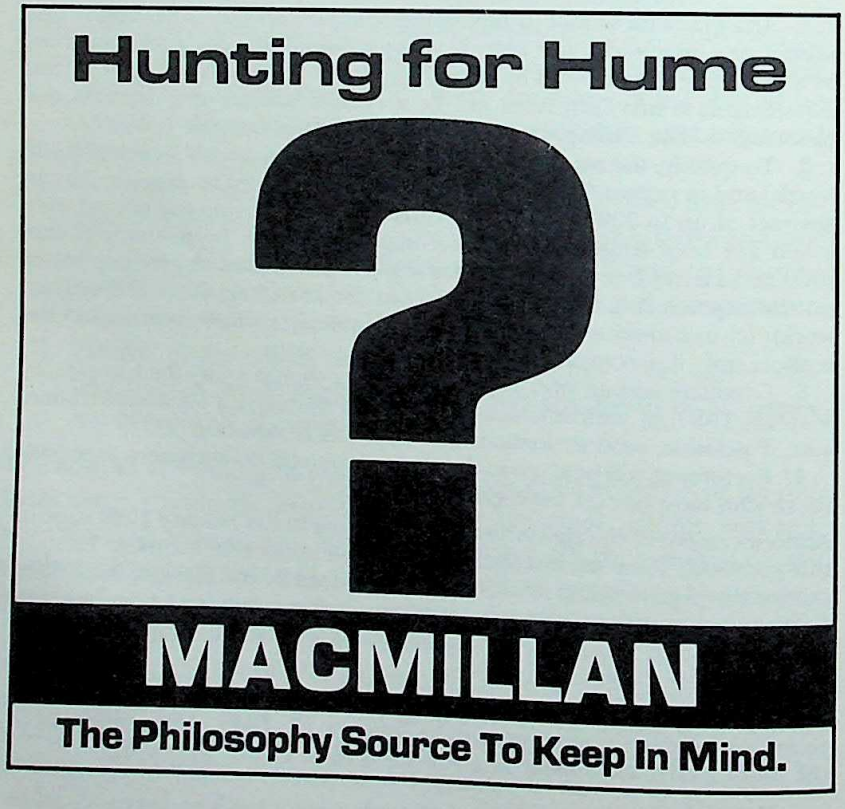
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

STATEMENT OF COPYRIGHT POLICY

The copyright law of 1976 grants certain new rights to libraries and educators who need to photocopy and certain new protections to copyright owners. In general, photocopies in excess of "fair use" copying cannot be made without authorization by the copyright owner. What constitutes "fair use" is not yet entirely clear, but some single-copying for personal use is permitted, and multiple-copying for classroom use is clearly limited: the article copied must be *short* (less than 2,500 words) or only an excerpt copied; the copying must be *spontaneous* (at the instance of the individual instructor, and needed so promptly as not to allow time to obtain permission); and *each copy must include a notice of copyright.*

The *Journal* authorizes copying *beyond fair use* under the following terms: the per-copy price for personal or educational copying is 10¢/page, with a minimum of 50¢, unless some special arrangement is requested by the author; the per-copy price for multiple-copying for classroom use is 5¢/page, where an instructor or the copy center involved applies directly to us for our permission.



CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

UNIVERSITY PRESS OF
New England

ON BEING MORAL . . .

Why You Should

The Pragmatics of Deontic Speech

JAMES W. FORRESTER

❦ ❦ ❦

A philosopher systematically considers rules governing the effective use of moral language in everyday speech, in the law, and in other contexts. He proposes 18 maxims to determine the effectiveness of using deontic speech—that is, in saying people may or should act in certain ways, and discusses the relation between the deontic and the moral to answer the question "Why should a person be moral?" He argues that following these maxims avoids many ethical Catch-22's occurring in life. *$28.00*

UNIVERSITY PRESS OF NEW ENGLAND
17½ Lebanon Street, Hanover, NH 03755

"A great and beautiful work of genius,"* now in a superb one-volume abridgement

Mind: An Essay on Human Feeling

Abridged Edition

Susanne K. Langer
abridged by Gary Van Den Heuvel
with a foreword by Arthur C. Danto

"One of the most audacious philosophical visions of recent times."
—Arthur C. Danto

The crowning achievement of one of our most seminal and refreshing philosophers, Langer's masterwork is now republished in an abridged edition that skillfully preserves the range, complexity, style, and coherence of the original three-volume *Essay.* Drawing on art and aesthetics as well as on psychology, linguistics, and the laboratory sciences, Langer frames a theory of mind that anticipates much of the work being done today in cognitive studies.

Chicago Daily News

$14.95 *paperback* **$45.00** *hardcover*

THE JOHNS HOPKINS UNIVERSITY PRESS
701 West 40th Street, Suite 275, Baltimore, Maryland 21211
or call 1-800-537-JHUP

CC-0. In Public Domain. Sarabai Nangia Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Cambridge University Press

The Origins of Democratic Thinking

The Invention of Politics in Classical Athens

Cynthia Farrar

Shows how the development of political theory accompanied the growth of democracy in Athens in the fifth century BC by analyzing the writings of the sophist Protagoras, the historian Thucydides, and the cosmologist Democritus in the context of the political developments and speculation about the universe. *$37.50*

On Religion

Speeches to Its Cultured Despisers

Friedrich Schleiermacher

Richard Crouter, Translator

Schleiermacher's *On Religion* is a central document of early nineteenth-century thinking about religion. This new translation of the original 1799 edition includes detailed annotation and a substantial introduction.

Texts in German Philosophy *$37.50 Hardcover* *$12.95 Paper*

On Language

The Diversity of Human Language-Structure and Its Influence on the Mental Development of Mankind

Wilhelm von Humboldt

Peter Heath, Translator

Hans Aarsleff, Introduction

This is an entirely new translation of one of the fundamental works in the development of the study of language. Originally published in 1836, it presents a survey of a great many languages and explores ways in which their various grammatical structures make them suitable vehicles of thought and cultural development.

Texts in German Philosophy *$49.50 Hardcover* *$14.95 Paper*

Now in paperback...

The Cambridge History of Later Medieval Philosophy

From the Renaissance of Aristotle to the Disintegration of Scholasticism 1100-1600

Editors: Norman Kretzman, Anthony Kenny, and Jan Pinborg

"...a well-documented and plausible new synthesis...it elegantly explicates many intricate philosophical arguments; it identifies promising avenues for further research; and it provides an extremely useful instrument for reference."—The Philosophical Review *$27.95*

The Cambridge Translations of Medieval Philosophical Texts

Volume 1: Logic and the Philosophy of Language

Editors: Norman Kretzman and Eleonore Stump

The first of a three volume anthology, this book is concerned with logic and the philosophy of language, and comprises fifteen important texts on questions of meaning and inference that formed the basis of Medieval philosophy. *$59.50 Hardcover* *$19.95 Paper*

At bookstores or order from

Cambridge University Press

32 East 57th Street, New York, NY 10022. Cambridge toll-free numbers for orders only: 800-872-7423, outside NY State. 800-227-0247, NY State only. MasterCard and Visa accepted.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVI, NO. 2, FEBRUARY 1989

2397129

REALISM WITHOUT INTERNALISM: A CRITIQUE OF SEARLE ON INTENTIONALITY*

IN his *Intentionality,* John Searle[1] provides a theoretical account of intentionality which depends on two principal theses: first, the content that belongs to intentional states is not constituted by anything that is not internal to the agent who possesses them (*internalism*); and, second, "people do really have them" (*realism,* or as he sometimes calls it, "intrinsicalism"). Most of the detailed claims and arguments in the book are advanced in support of or drawn as consequences of one or other of these two theses. In a short discussion, I am bound to omit most of these details, some though by no means all of which are terminological variants of points well established in the literature; but let me begin with one or two crucial points of detail so as to make the larger theses more perspicuous.

All intentional states are said to consist in an intentional content in a psychological mode. A belief, for instance, is in a different psychological mode from a desire, and each of these is in a different mode from an intention, which is not reducible to either or both. The contents are explicated in terms of what seems to be a generalization of the notion of truth conditions. Thus, Searle introduces the notion of satisfaction conditions. The satisfaction conditions of a belief are its truth conditions, whereas those of a desire are the conditions under which it is fulfilled; and those of an intention, the conditions under which it is carried out.

For reasons that are not made fully explicit, the intentional states involved in perception (he calls them "visual experiences") and in-

* I am indebted to Marcia Cavell, Donald Davidson, Josh Guttman, Sidney Morgenbesser, Carol Rovane, John Searle, Claudine Verheggen, Stephen White, and the Philosophy of Language and Mind reading group at Columbia University for helpful discussions on the themes of this book.

[1] New York: Cambridge, 1983.

0022-362X/89/8602/57–72

© 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tentional action (intentions) are supposed to be basic to us as the sort of creatures we are. There are lengthy chapters spelling out what goes into the contents of perceptions and intentions. I shall restrict discussion to the former.[2] Perceptual episodes have intentional content and these are, in the usual way, specifiable in a that-clause: "I have a visual experience that there is a yellow station wagon there." Such simple specifications, though perhaps all right for other sorts of belief, are insufficient for perceptual beliefs. The satisfaction conditions of these must contain a complication; they must include the condition that the visual experience "must itself be caused by the rest of the conditions of satisfaction of that visual experience." So we get contents of the following sort: I have a visual experience that *there is a yellow station wagon there and that there is a yellow station wagon there is causing this visual experience.* This is described as the causal self-referentiality of perception.

Why is this complication introduced? For even if it is phenomenologically intuitive that perception involves such a causal link between the experience and a state of affairs, it is hardly obvious that this warrants the self-referential element in a specification of perceptual *content.* Searle nowhere answers this question directly, so one has to turn to the work to which it is put in the overall doctrine. And that work is hard and multifarious and spans four chapters of the book. Yet there is a common thread in all its uses. In the end, I suspect, the real motivation lies in the first of the two theses I mentioned at the outset, the internalism.

For one thing, he observes that internalist theories have often been charged with conceiving of thoughts as wholly general and not taking in particular things. Thus, for them, the content of a perceptual thought can remain the same if a quite different yellow station wagon is present. This is a charge he thinks worth repudiating, for the content should take in the particularity—one should be able to specify thoughts about particular station wagons. Those who usually lodge the complaint against internalism assume, according to Searle, that there can be no response that does not give up on or add to the internalism. They think only externalist conceptions of content which appeal to external causes of content will solve the "particularity problem." He brings this out, as they often do, with examples of twin agents on earth and twin earth. So, for instance, the one on earth sees his wife Sally, the one on twin earth sees his wife twin Sally.

[2] The special feature of self-referentiality which Searle thinks attaches to perception and intention has already been anticipated for intention by Gilbert Harman [see his "Practical Reasoning," *Review of Metaphysics,* XXIX (1976): 431–463] and has received a fair amount of discussion in the literature.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Externalists achieve the particularity by insisting that what determines the content is that Sally causes the perception of the agent on earth, and twin Sally the perception of the twin agent. Searle objects that this solution is from a third-person point of view, how an observer tells which one is being perceived. But he thinks an internalist is committed to a first-person solution to the particularity problem. The question, then, must be: What is it about *my* experience that requires that it be satisfied by the presence of Sally and not by any woman with various features type-identical with her? There is no objection to an answer appealing to causality, so long as there is no concession to externalism. Thus, no *external* causes are constitutive of content, even though we may be thinking of external things. This is why he says that the causality must be *part of* the intentional content. A fairly elaborate apparatus is set up to do this, but at its center is the self-referential element. The idea is that it is part of the intentional content of my perceptual thought that Sally stands there, that Sally's standing there is causing me to have that perception.

This allows that Sally in fact not be there. All that would mean is that the satisfaction conditions that are specified in the content have not obtained. But whether they obtain or not, whether one is veridically perceiving or hallucinating, the content is the same. Yet my twin's content is not the same, since it is part of his intentional content that it be *his* wife, twin Sally, who is causing him to have the experience that Sally is there. And so, even if twin agents have phenomenologically identical experiences, the contents are different. The causal self-referentiality clause in the specifications sees to that. The appeal to external causes is, therefore, unnecessary.

It is thus the governing internalism that motivates the feature of self-referentiality in perceptual content; and the feature is exploited toward various ends in the philosophy of language, ends which tie in quite naturally with the internalism about intentional states—to argue against the causal theory of reference and thus the idea that meanings "ain't in the head," to deny the existence of *de re* thought, and to offer a Fregean account of indexical expressions. Thus, the "particularity problem" about content is carried over to the question: What is it that makes Jones *refer* to Sally rather than to twin Sally when he says "Sally!" Causal theorists like Saul Kripke are scolded for giving an externalist, third-personal answer; by situating the issue in the context of a speaker's intentional contents (instead of resting, as in his earlier work, with talk of "descriptions") and by introducing the self-referential element in these contents, Searle answers the question without compromising the internalism. Also, since it is particularity, Searle says, that prompts philosophers to

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

think that there is a category of intentional states which take objects as part of their content (*de re* thought), then that category is dispensable, if particularity can be achieved along the lines he has suggested. And, if one thought that indexicality cannot be so easily handled, since, unlike *de re* thought, indexical thinking can hardly be denied to exist, Searle accepts this difference but once again invokes the same idea: when on a particular occasion someone speaks a sentence containing an indexical expression (or has an indexical thought), the content of that thought reveals the relations that the object he is referring to has to the very utterance that expresses it. So, to the standard demand that indexical utterances require a completing sense, Searle thinks it is enough to respond by saying that, once we see intentional content as containing the self–referential element, the completing sense is right there in the content specified with that additional clause.

I shall not raise a question about Searle's repudiation of *de re* thought or about his analysis of indexicals. On the first matter, despite some questions about his treatment of Tyler Burge, I am inclined to agree with him; on the second, though his analysis ignores the epistemological issues surrounding the question, it certainly achieves the restricted semantic task it sets itself.[3] I do, however, have some disquiets about the internalism that underlies these more specific of his conclusions.

What, one should ask at the outset, is the motivation for internalism? Searle is nowhere explicit about this, as others have been. In the mountain of writing on the subject (much of it before the publication of this book), three motives seem to be most prominent. The first simply finds in internalism a *metaphysical* intuition about the mind. In the writing it is usually expressed as an intuition about how intentional facts must supervene on facts only about the interior of agents.[4] The second is the thought that, if a *science* of the mind is to

[3] It should be clear, and Searle will surely admit, that it can be achieved in quite other ways than by the introduction of the causal self-referential clause.

[4] It is not clear to me that there is an intuition here. Many deny having it. In any case, it certainly is not a prephilosophical one. (I think there is a lot of evidence that the intuition seems to vary with whether those who have it have been educated in a philosophical climate dominated by admirers of Wittgenstein or, say, of Thompson Clarke.) It is often said that the intuition is most vivid in the imagined case of the brain-in-a-vat. I do not see that it is more or less so. It may be that what is intuitive here is that the brain's utterances and, so, its thoughts are the same as someone else who is uttering the same noises in another environment; but the basis of this intuition may be *not* that the two subjects have the same *internal* makeup, but rather that their utterances are best correlated with the same *external* environment, even if this requires thinking of the mad scientist manipulating the brain as part of

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

emerge, then any description of the mind had better leave out the relations in which agents stand to external things.[5] The third arises from the worry that, if external things were allowed to constitute the contents of intentional states, then these states would no longer be fit for their role even in the common-sense explanation of human actions. This is because a focus on external things would leave out agents' *cognitive* worlds, their *conceptions* of things, and it is these latter which account for their actions.[6]

Searle does not acknowledge these points in the literature and though, in a few places, he says things that suggest that he is aware of and endorses some of them, their occurrence is too casual and too buried in the text to impress one as being central to his commitment to internalism. For instance, he says repeatedly, "Of course meanings are in the head, where else are they going to be?," suggesting that he has in mind the first motivation. But this is too rhetorical to say anything very specific. It is compatible with a reading that takes meanings (or intentional contents) to be *token-identical* with inner states of agents. And that is by no means the same as internalism in the sense which the book advances which is better characterized, as I said, in terms of the supervenience thesis. Token identity of states

---

the brain's perceptual mechanism. (Where there is no manipulator, it may be that one should assume and look for some other nonstandard perceptual and learning mechanism.) The nonstandardness of this cannot be a source of dissatisfaction, since it only matches the nonstandardness of the imagined case of such a brain.

The intuition is sometimes described as Cartesian. It is possible perhaps to read it into the first of Descartes's *Meditations*. But, in the second, Descartes emphasizes much more the authority an agent has upon his own mind and its contents. This would constitute a quite different motivation for internalism than the intuition I have mentioned, which can stand (and in recent writing has stood) independently of considerations of first-person authority. Searle's overall position does indeed rely on a claim that internalism alone will capture such authority (see my discussion of his attack on the indeterminacy thesis) and perhaps therefore is more appropriately describable as Cartesian. It is not even clear to me that, if one wishes to motivate internalism by the idea that the contents of one's mind should be characterized in a way that allows for the coherence and possibility (however remote) of Meditation I's skepticism about the external world (see fn. 20 for more on this motivation), one would have to embrace the second Meditation II's stress on first-person authority. This is especially so if that skepticism is generated by more modern thought experiments such as the possibility of one's being a brain-in-a-vat. Donald Davidson has persuaded me that I should be less confident of this separation of issues and motivations in Descartes's own procedure.

[5] For a forceful statement of this view, see Jerry Fodor's "Methodological Solipsism as a Research Strategy in Cognitive Psychology," *Behavioral and Brain Sciences*, III (1981): 63–73.

[6] Brian Loar argues along these lines with great clarity in his "Social Content and Psychological Content," in R. Grimm and D. Merrill, eds., *Contents of Thought* (Tucson: Arizona UP, 1985). It is a line of thought which goes back, of course, to Frege's arguments for introducing a notion of sense.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

possessing intentional content with states of the brain is a thesis that is fully compatible with externalism.[7] At other places, Searle says that his internalism merely extends Frege's notion of sense to the study of intentionality (245), suggesting the third motivation. But here again nothing is said to indicate that the Fregean reasons deriving from considerations of cognitive content force the internalism. Perhaps he thinks these standard motives are too deep within our sensibilities to need stressing. But I suggest that his fully thought-out motivation for internalism lies elsewhere. I shall return to it later.

I raise the matter now, because it will seem initially hard to the reader how controversial to take Searle's internalism to be, at a time when he is in such numerous company. Many, if not most, who are externalists have insisted, for one or other of the reasons I have just mentioned, that a *second* notion of content is required which *will* be more purely internal. Thus, for instance, Hilary Putnam,[8] whom Searle attacks, has argued that we need another notion of meaning than one tied to world-involving concepts like reference and truth, a notion defined instead in terms of the concept of verification. (Although the point here is made about meaning, I am assuming, with Putnam and everyone else, that it carries over to intentional content.[9]) Even Kripke, whose idea of reference Searle attacks but who has made no such concession to a second more internalist notion of content, nevertheless has raised a "puzzle about belief," a puzzle which clearly only arises if one embraces an externalist account of content based on that idea of reference. And others broadly sympathetic to Kripke's idea have explicitly taken the puzzle to force such a second notion.[10] So, in the face of these rather major concessions to

[7] For a convincing account of this compatibility, see Davidson's "Knowing One's Own Mind," *Proceedings and Addresses of the American Philosophical Association,* LX (1987): 441–458.

[8] See Hilary Putnam, "Reference and Understanding," in his *Meaning and the Moral Sciences* (London: Routledge & Kegan Paul, 1978). For a more explicit connecting of verificationism (in particular, Michael Dummett's verificationism) with internalism, see Colin McGinn's "Realism and Content Ascription," *Synthese,* LII (1982): 113–134; and William G. Lycan, *Logical Form in Natural Language* (Cambridge: MIT, 1984), ch. 10. I discuss the implausibility of making this connection in my "Meaning, Holism and Use," Ernest Lepore, ed., *Truth and Interpretation: Perspectives in the Philosophy of Donald Davidson* (New York: Blackwell, 1986).

[9] The precise ways in which the connection between meaning and content must be spelled out is a delicate matter, but throughout this discussion I shall assume that, however it is spelled out, these disputes about internalism and externalism apply to both. Searle himself takes a rather strong view of the connection, placing intentionality as conceptually prior to meaning in a very strict sense (see 26–29). One does not have to take such a view to make the assumption.

[10] See Loar, *op. cit.*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

internalism by his opponents, it is hard to assess how much bite Searle's own internalism can have.

Of course, one can expect that it will have teeth against externalists who work exclusively with a *single* notion of externalist content. But, even here, I think Searle's failure to stress the standard motivations for internalism must have blinded him to the subtleties of such externalist positions as one finds in, say, Gareth Evans and Burge. These externalists are careful to speak to some of these motivations, either accommodating them somehow without compromising the externalism or systematically repudiating them. Evans accepts the third Fregean motivation I mentioned above (see fn. 6) and argues to accomodate it in his externalism by introducing a notion of *de re senses*. Burge offers an externalist position, while arguing in depth against the first two motivations. Thus, Searle's entire attack in the last third of the book ignores the strongest version of the thesis he is attacking.[11]

I have so far been stressing externalism's sensitivity to standard internalist yearnings, something not to be found in Searle. But Searle may well wish to be arguing that externalism is false or unnecessary, no matter how accomodating it is toward these yearnings. If this is so, one should expect that he is sensitive to *their* motivations in return, arguing against them in detail. But there is not much evidence of this. To be fair, there are many different versions of externalism formulated with quite different goals in mind, many of them not explicitly stated in the literature. Even so, judging from the absolutely key role played by the idea of the causal self-referential element in content, it is hard to avoid the conclusion that Searle has approached externalism with a very limited conception of its aspirations as a doctrine about content. Let me explain.

Sometimes, and especially in its early phase, the issue between externalism and internalism was expressed as one about diverging commitments to and against an indexical element in certain contents. Twin-earth examples were introduced, for instance, with a view to finding a hidden indexical element in our thoughts about

[11] Evans, *Varieties of Reference* (New York: Oxford, 1982), ch. 1; and Burge, "Individualism and the Mental," *Midwest Studies in Philosophy*, IV (1979): 73–121. Evans's externalism has been subtly elaborated by John McDowell in his "*De Re* Senses," in Crispin Wright, ed., *Frege: Tradition and Influence* (New York: Blackwell, 1984), but this appeared after the publication of Searle's book. In my "An Externalist Account of Psychological Content," *Philosophical Topics*, XV, 1 (Spring 1987): 191–226, I offer an externalist account that accommodates the Fregean elements and also avoids any commitment to the *de re* or Russelian propositions. I entirely accept Burge's criticisms of the first two motivations for supervenience and internalism, but in the same paper I criticize his version of externalism also.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

certain natural kinds, which an internalist conception of thought would (apparently) fail to capture.[12] The generalization of this to all perceptual and other sorts of thought is what Searle calls the challenge of "particularity." It is this challenge, as we saw, which he set out to answer with his apparatus of causal self-referentiality. And one should say that, even if it is not the only apparatus that would meet the challenge, it certainly meets it quite adequately. In doing so, he has shown the initial emphasis on indexicality in these disputes to be misguided. All the same, it is by no means the case that this is the hardest or most interesting challenge that the externalist throws down, and Searle makes things easy for himself by concentrating on it exclusively.[13]

The fact is that there are quite other motivations for the doctrine that Searle ignores. What are these? The most convincing motivation it seems to me is this. If one believes that thought and meaning must be public phenomena (and I shall assume it here without question until fn. 20, which is crucial to my overall argument), then the following is a good question: How shall we characterize thought and meaning such that its public availability is ensured? I think an externalist characterization alone will satisfactorily ensure it. If another's meanings and propositional attitudes are determined by items in a world external to her, then it is neither surprising nor avoidable that they are available to one who lives in the shared environment.[14]

I cannot possibly establish in this discussion that this is the only satisfactory answer to our question, but let me say something brief against two quite different answers (which seem to avoid externalism) so as not to make it appear as sheer prejudice. One answer, oddly enough, is given by John McDowell,[15] who is an externalist, but the answer seems to be independent of his externalism. McDowell argues

[12] See especially Putnam's "The Meaning of Meaning," *Mind, Language and Reality* (Cambridge: Harvard, 1975).

[13] Indexicality really should come in at a quite different place in the externalist doctrine. If externalism were true, then the points of connection between the external world and the contents of agents' minds would very plausibly occur at the point of agents' indexical contents. Read this way, particularity is an essential part of the characterization of an externalist position without being its motivation.

[14] This answer must, of course, take into account the fact that many contents are very far removed from these external elements, that is, much more mediated by theory. Also, I do not think that an externalist motivated in this way is in any way committed to saying that every indexical or perceptual content must have an object or event as external cause. I discuss this and other details of the externalist method in my "Externalist Account of Psychological Content"; see especially section III and the criticism of Evans's similarly motivated externalism in section IV.

[15] "Anti-Realism and the Epistemology of Understanding," in H. Parrett and J. Bouvresse, eds., *Meaning and Understanding* (Berlin: de Gruyter, 1981).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

that skepticism about other minds can only be answered if one takes the right view of the epistemology of understanding others. In particular, we must see understanding as a form of *direct* perception of another's meanings and thoughts in their speech and action. To such a view, the idea that these meanings and thoughts are theoretical posits, constructed partly out of the relations in which they stand to the environments of their possessors, will presumably seem quite false. At the very least, their *direct* availability to perception will make the more roundabout idea unnecessary. Even if externalism is true, it will be irrelevant to the question about public availability. In my view, this naive realism about others' thoughts is perfectly all right as a piece of descriptive phenomenology, since it is usually the case that our understanding of another is noninferential. But, epistemologically speaking, it is beset with an old, and to my mind insoluble, problem that attaches to naive realism about the perception of anything whatsoever, viz. that there is no satisfying account of perceptual error—in our case, of the misunderstanding of another's meanings or thoughts. All efforts by naive realists to deal with this problem have not been very compelling. The second alternative answer is this. Someone inclined to think that there is a reduction of meaning and content via certain internalist versions of functionalist doctrine may hold that our question is perfectly and easily answered, since the items such a doctrine appeals to are clearly public states (peripheral stimuli, neural states, bodily motions). The trouble, however, is that these (crucially, the first two items) are public in a way that is irrelevant to the spirit in which the question was asked. When one talks of the publicness of language and mind, one means their availability *literally* to a public and not merely to those who, armed with relevant instruments and with a reductionist theory (yet to be forged), can examine these items in an agent.[16]

Now, as it happens, the externalism forced by having to give a

[16] A proper appreciation of why neither McDowell's idea nor such a functionalism answers our question will show that publicness is never secured by adopting the stance of simply saying that our thoughts are available to others via our behavior. If the availability is not additionally routed through the element external even to our behavior (that is to say, external even to what carries our behavior, our bodily motions), then the stance will inevitably have to rely on one or other of these two unsatisfactory answers.

Platonism may be thought to provide a third alternative answer to our question, an answer that is noninternalist and yet not externalist in any thing like the sense on which I have insisted. But, so far as one can tell, Platonism is merely an assertion of the objectivity of meaning and content. That only amounts to saying that, if two subjects believe or mean the same thing, then there is something objective that they both believe or mean. In itself that does not say on what basis they understand each other or others understand them, so it does not so much as address our question.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

satisfactory answer to this question about publicness does not (or need not) by any means amount to the externalism involving what Putnam and others in their twin-earth and other such thought experiments call *wide* content. That is, it need not get its point or rationale by a scientific essentialism or by Burgean notions of social, linguistic norms and practices that are central to the thought experiments that give rise to the idea of wide content. These latter externalisms have been formulated with quite other goals than that of assuring publicness; but it is not at all obvious that all these goals are good ones or that they cannot be achieved without a commitment to wide content, and it is therefore not obvious that wide content is a necessary feature of propositional attitudes. Certainly, as I have been saying, a general commitment to externalism does not require it. Narrow content may be all that is necessary so long as it is publicly available (thanks to its externalist constitution). This may seem initially startling, since narrow content is always taken to be an internalist notion. There are good reasons, however, to doubt that the narrow/wide distinction coincides with the internalist/externalist one. Taking narrow content to be defined in contrast with wide content requires a contrast with a *very specific* externalist notion of content derived from the views of Putnam, Kripke, or Burge. The need for narrow content arises for those who accept a notion of (wide) content derived from these views, because the latter notion fails to capture agents' cognitive worlds. I had earlier presented this as the third motivation for an internalist notion of content, since I was reporting on the current ways of thinking on the subject. I am now denying that this motivation has to be fulfilled by an internalist notion. If this is right, then narrow content which captures agents' cognitive worlds can be externalist. Further, if it is also correct, as I am arguing, that there is no need for wide content, then we may drop the word 'narrow', since it has lost its contrast.

The arguments philosophers have given for introducing wide content into externalism—such as that it accounts for how one might gain knowledge of the world by attributing it to others, or that it affords one a distinction between theory change and meaning change, or that it alone captures the normativity of meanings, or quite simply that it is what we attribute to others in everyday speech —need careful attention even if one does not find them convincing.[17] Searle does not devote any energy to this task. Thus, although his arguments against Putnam and Kripke's externalist causal-theo-

[17] In my "An Externalist Account of Psychological Content," I explicitly argue against the idea of "wide" content (and speak to its motivations) while defending externalism.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

retic views are in many ways trenchant, his case would have been much stronger had he taken up these motivations in detail.

But the crucial lack, at least from my point of view (since the externalism I favor does not amount to wide content in the sense that Searle attacks when he attacks Putnam and Kripke) is his failure to worry at all about the publicness of intentional content and whether and how internalism can allow for it. Does he find the question about publicness which we posed earlier a worthwhile one? If not, why not? If so, does he embrace McDowell's answer on the question? (Since he has, famously, opposed functionalism, the other answer is presumably not available to him). Until he confronts these issues, his internalism will remain precarious.

I began these comments by saying that Searle does not make much of the standard arguments for internalism. This suggests that his first principal thesis is prompted by some more underlying unease. What is fundamentally wrong with externalism, according to Searle, is a commitment to a third-person point of view. It leaves out how things are for the agent, the first-person point of view. And he thinks that, unless one characterizes content from the agent's point of view, one is refusing to treat intentional states that possess it as being *intrinsic* to agents. All one is left with is a *stance* or a metaphorical way of talking; one is not attributing *real* states to agents. For, if one sees intentional contents as constituted by what another (an interpreter) attributes to an agent, then the interests of the interpreter enter into the attributions, reducing them to a merely convenient and instrumental way of talking about agents' behavior. Not only does externalism get things the wrong way round, it makes it impossible to take a realist attitude toward intentional states. This has brought us to the second principal thesis in the book.

A recent example given by Daniel Dennett[18] may help clarify this. (It is explicitly directed against Searle's intrinsicalism or realism.) Take a machine in New York City which accepts quarters and hands out chocolates. As far as the machine is concerned, the Panamanian balboa will do just as well—its physical shape and contours are acceptable and exchangeable for chocolates. One can imagine, however, that, while in New York City and in the charge of some local owner, only quarters are acceptable, the exchange upon being fed a balboa would count as a *mistake*. In Panama ("the poor man's twin earth," as he calls it), under a quite different charge, things would be just the other way round. Although, of course, no one will think *this*

[18] "Evolution, Error and Intentionality," in his *The Intentional Stance* (Cambridge: MIT, 1987).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

machine has intentional states, an instrumentalist is supposed to take the view that human beings are just like this machine in crucial respects. What counts as an intentional state with one content rather than another is a matter of the social and interpretive context in which the human beings (machines) are lodged. One sort of thought or utterance might count as a mistake in one such context, another in another. This is an instrumentalism that Dennett has often espoused and, while doing so, has often listed several prominent philosophers of mind as being on his side. Searle and other realists reject the idea that the situation with human beings (where intentionality genuinely has application) is at all like it is with the machine. For them, our talk of this machine *accepting* coins or making *mistakes* is mere talk, not only because the machine's abilities are very limited, but also crucially because intentional states, they say, are *not* up to the interpreter and to the social context in which agents (to whom they are attributed) live; they are *intrinsic* to agents. The only way to get a true characterization of them, therefore, is to respect this intrinsicness, i.e., *their* point of view.

This way of drawing the antagonism between realism and instrumentalism about intentional states is, it seems to me, much too simple. In fact, it had better be so, since neither Dennett's nor Searle's position seems very attractive. What complicates things is the fact that taking an interpretive, or what Searle calls a third-person point of view, need not by any means have the consequence that the intentional contents attributed to human beings will be interest-relative in anything like the sense suggested by Dennett's analogy with the chocolate machine. There are surely constraints one may place on attributions by a third person which do not leave out the agent's point of view. Take, for instance, the early disputes over the nature of radical interpretation, where it was thought that meanings and beliefs were to be attributed to an agent by an interpreter with the constraint that overall agreement between agent and interpreter be maximized. It was justly protested against this view (though it is not clear that anybody really held it) that this would leave out the *agent's* point of view and thus the agent would be said to have propositional attitudes she did not *really* have; so a quite different constraint was proposed which sought not to maximize agreement, but to minimize unexplained error. Now, in this last sentence, I have raised a genuine question about realism regarding intentional states and the first–person point of view. And the question is raised *within* a third-person characterization of intentionality, i.e., within the context of an interpreter's attributions. This suggests that, though *some* version of the opposition between the first- and third-person points of view is

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

relevant to the dispute between realism and instrumentalism about intentionality, it is not the version we find in the dispute between Dennett and Searle. The version that is relevant turns not upon intrinsicness in the sense of internalism and the Cartesian perspective, but rather on what constraints to place on attributions of content to an agent, even within, if need be, a third-person point of view. Only *some* constraints will lead to attributions that capture the point of view of the agent to whom the contents are being attributed. That is what realism in this area is about.[19]

None of this, at least in its general form, should be surprising. The commitment to a first-person point of view, after all, cannot and does not (even for Searle) amount to so strict a Cartesianism that it surrenders the publicness of meaning and content.[20] If so, a third-

[19] It should be obvious that this is only a necessary and not a sufficient condition for realism about intentional states. Other necessary conditions might be a certain holistic complexity including a self-reflexivity regarding intentional states on the part of the agent, which of course is why talk of the vending machine as having such states is instrumental; and also certain *limitations* on the extent of indeterminacy to which the attributions are subject. Imposing the right constraints goes a long way toward reducing indeterminacy. Of course, it will not eliminate it altogether. But see the ensuing discussion on the indeterminacy that remains.

[20] So far I have been writing for a reader who agrees to the publicness of meaning and content and would wish to *ensure* it in the characterization of these concepts. Since Searle nowhere denies publicness, I have assumed that he, too, is such a reader. Even so, this may be the place to turn to another sort of reader and address some underlying epistemological concerns.

Assume the following response by someone, possibly Searle: publicness is only a contingent aspect of meaning. Thus, although there is not any great need to deny that we *in fact* discover each other's meanings and thoughts along the lines I am suggesting, that does not mean that these external elements *constitute* meaning and thought. One's thoughts would be just what they are even if there were no external world and, in that case, they would *not* be discoverable. That is a coherent conceptual possibility.

So let me turn now to saying something *directly* in defense of externalism rather than defend it, as I have, by introducing it as the only grounding for publicness. Instead of looking to attributions of content to one agent by another, one must now look to one's own specifications of content. One can think of Searle's own specifications, "There is a yellow station wagon in front of me"; or Descartes's, "I am sitting by the fire in my dressing gown." Now a question arises as to what right internalists like Searle or the Descartes of the "First Meditation" have to concepts of objective and external things such as station wagons, dressing gowns, and fires in the specifications of their thought or experience. Whence this elaborate conceptual structure?

Internalists of a somewhat different stripe, such as Hume or A. J. Ayer, have, as is well known, honestly tried to deal with this question by trying to show that these concepts are derived from or constructed upon genuinely simpler inner objects: sense impressions or sense data. One may assume that the unworkability of the phenomenalist program makes their answer unacceptable. One may even safely assume that Searle finds it unacceptable. An alternative answer is given by the externalist: our experience (and thought) is specified this way because much of the time it is the experience of objective and external things. But what answer can Searle give? Clearly, it would not be enough to say that we gain this conceptual

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

person point of view should be compatible with a commitment to it, i.e., since an agent's thoughts are discoverable by a public, there cannot be any wholesale wrongdoing or thinking in taking a third-person point of view on an agent's thoughts. If this is so, there cannot be anything to Searle's more recent attack on Quine and Davidson's commitment to the indeterminacy of meaning and content.[21] He argues that indeterminacy can be avoided if one shuns the third-person point of view of radical translation and interpretation. This brings an unnecessary opposition between a third–person point of view and first-person authority. If what an agent believes and means is publicly discoverable, then a radical interpreter working with the right constraints may presumably discover them. And, if a radical interpreter's discoveries here are ineradicably subject to indeterminacy, then thumping the table with the authority an agent has over his own thoughts and meanings will not eradicate it at all. What one should conclude, instead, is that indeterminacy is, in the end, harmless and leaves unthreatened the notions of meaning and content over which we have first–person authority. Such authority is undeniable, but it does not have the significance Searle sees in it.[22]

---

structure by the having of experience, because something will have to be said about what about or in the experience provides it. The internalist–phenomenalist, and externalist have said something about it, but what can Searle say? It would be utterly implausible to suggest that this entire panoply of concepts is innate. (Some conceptual structure is no doubt innate, but that is compatible with an externalist answer to the question I have posed.) So far as I can see, there is no plausible alternative to the externalist answer.

This may not be absolutely conclusive, but it does seem to pose an unanswered challenge to internalism, and it would be irresponsible and complacent of internalists not to face it squarely. Until it is answered, the scales are visibly tipped in favor of the externalist. Moreover, it is only philosophical questions and considerations such as these that *will* tip the scales one way or another in a dispute, which, while debated at the level of intuitions (as in fn. 4), will always seem to us to be a standoff.

[21] See especially Searle's "Indeterminacy and the First Person," this JOURNAL, LXXXIV, 3 (1987): 123–146.

[22] It should also be noted that the argument that only the first–person point of view will allow for realism about intentional states is quite independent of the use of the idea of a first-person point of view as a special authority for which an interpretive or third-personal perspective will not allow. It is not so clear, however, that Searle realizes this. One may accept the criticism of the instrumentalist position Dennett seems to take on the ground that it leaves out the first-person perspective, without in any way embracing the rest of Searle's idea of what goes into the first-person perspective, i.e., a sort of authority the possessor of intentional contents has which will get rid of all indeterminacy. The term 'first-person perspective' is doing too many different things for Searle and there is no essential connection between them. See fn. 4 for a further disentangling of the idea of this Cartesian first-person authority from other motivations for internalism such as the supervenience thesis and the coherence of skepticism about the external world. None of this disentangling should give the impression that I deny that we have authority over our own

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

I have been trying to extricate the idea of a distorting interest relativity from the general idea of a third-person point of view which has misled Searle and others (including Dennett) to associate the latter with an antirealism about intentionality. I have also tried to dissociate the third-person point of view from any abandoning of first-person authority. But, having done so, I should point out that there is something further which has misled Searle and other philosophers in the discussion surrounding the idea of a third-person point of view. Many have been put off by the idea that a focus on the process by which an interpreter (a third person) discovers content can illuminate *what is discovered* about an agent; the former, being of epistemological interest only, cannot give us the nature of the states themselves. But, if my dialectic is right, this puts the emphasis in the wrong place. It is not that interpretation constitutes content. Rather, it is because content is externally determined that it is a public phenomenon. And, because it is a public phenomenon, interpretation and the constraints we put on it will help shed light on intentionality.

In the very last chapter of the book, Searle introduces another ingredient in his realist thesis. To avoid instrumentalism about intentional content, one is not allowed to find it in subjects who lack the appropriate hardware; this is also part of his attack on functionalism. The idea of appropriateness here is notoriously unclear. Searle speaks vaguely of how the hardware must have the same "causal powers" as the brain, if it is to be a carrier of intentionality. This notion, it appears, can only avoid the mysteriousness many have

---

intentional contents. I entirely agree with Searle that we do, and I even agree that, if one is an externalist of the sort Burge and the early Putnam are, then there is some doubt that such authority can be retained. But those externalisms should not be equated with all third-personal approaches to the study of meaning and content.

In recent work, Searle has made things worse. He has added to his already overloaded use of "the first-person point of view" by introducing considerations having to do with "consciousness" and "qualitative states." Even here he tends not to keep separable things separate. In conversation, he has contested the claim that the idea of consciousness, as it occurs in "I am conscious of having the belief/desire that *p*," has a quite different point and use than when it occurs to characterize the special *felt* quality of qualitative states. But they are different since, even though we have authority over our own qualitative states, it is only the former occurrence that *says* something about self-knowledge and can be understood in terms of an iterated belief operator (and thus eventually from a third–person perspective); but such an operator is beside the point for the latter occurrence. And it is only the former occurrence that concerns the subject of his book. I shall not deny that the precise relations between these is a large subject which needs scrupulous handling. I am only complaining that not being careful about it is what allows Searle to run away with the impression that his attack on the third-person point of view is a monolithic argumentative strategy, which it is not and cannot be.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

found in it by embracing something equally implausible: a parochialism that others have found in it. So far as I know, Searle has nowhere, not even in subsequent publications, done very much to deal satisfactorily with this dilemma. It is also left unsettled here what the "hardware" criterion for realism has to do with the criterion involving the first-person point of view.

I have not been able to discuss or even mention a number of important and substantial things in this book, most regrettably his notion of a "background" to the possession of intentional content. There is also an elaborate discussion of how intentions cause actions, and a solution to the problem of deviant causal chains is proposed in the course of it. Throughout the book, Searle admirably and undistractedly focuses on the deepest and most fundamental issues in this area of the philosophy of mind.

Intentionality has, in many ways, been a pleasure to read. As I have argued, the theory of content it offers is as yet unfinished, and many questions remain to be answered, some of them quite basic. I hope I have conveyed, however, how interesting a work it is and also something of the fierce seriousness of Searle's attack on its themes. Many of us who have read him before have come to expect two things in all his writing: utter clarity and utter confidence. These are here, too. We are never in doubt about the meaning of any claim in the book. He is never in doubt about its truth.

AKEEL BILGRAMI

Columbia University

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

A PREFERENCE FOR SENSE AND REFERENCE*

THE main purpose of this paper is to examine a number of alternative accounts of the semantics of propositional-attitude reports, particularly English 'believes that' sentences. My aim is partly to promote the attitude expressed in the title of this paper and tentatively to offer a proposal in keeping with that attitude. But my aim is also to enhance our grasp of what a good explanation in the theory of meaning is by looking at the kinds of evidence and arguments which may be relevant to it.

In seeking an account of the semantics of attitude attributions, I seek to understand the workings of the actual sentences of natural language. I am not engaged in the kind of enterprise associated with Gottlob Frege and W. V. O. Quine of trying to construct a superior language for the serious purposes of science. My enterprise is the one Donald Davidson[1] distinguishes with the words: "the task of a theory of meaning as I conceive it, is not to change, improve or reform a language, but to describe and understand it" (*ibid.*).

What is essential to a semantic theory, in the first place, is that it accurately predict the observable semantic features of whole sentences. The form that such predictions take will depend upon the kind of theory with which one is working. A theorist working within a Davidsonian theory of meaning—and I am one—will aim to prove the familiar *T*-sentences: '*S* is true iff *p*', where *S* is a structure-revealing description of an object-language sentence, and *p* is a statement, in our metalanguage, of exactly the conditions under which that object-language sentence is true. What is crucial here is to get the *T*-sentences right. When we are in the fortunate position of constructing a theory of meaning for our own home language, the evidence is not hard to collect; it is often not hard to tell when a *T*-sentence is correct. The trick is to construct a theory that proves it.

The further evidential constraints to which one wants the theory to be answerable will depend upon exactly what is to count as a good description and explanation of the linguistic phenomena under scrutiny. A methodological conservative might seek only to articulate

* Earlier versions of this paper were presented at University of Pennsylvania, University of Wisconsin/Madison, and University of California/Berkeley: my thanks to those who participated in the discussions. I am also indebted to George Boolos, Sylvain Bromberger, Michael Byrd, Noam Chomsky, Berent Enc, Jim Higginbotham, Richard Larson, Peggy Speas, Dennis Stampe, Thomas Uebel, Bruce Vermazen, and Crispin Wright.

[1] "Truth and Meaning," *Synthese*, XVII (September 1967): 304–323.

0022-362X/89/8602/73–89 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

a body of knowledge, possession of which would suffice to interpret utterances made by speakers of the language. For the methodological conservative, two semantic theories that prove exactly the same *T*-sentences, but employ different axioms and mechanisms, are likely to be as good as each other. Each achieves what needs to be achieved, but they do so in different ways. There is, then, no need to seek further evidence that would count for one theory and against the other.[2]

From the point of view of the methodological liberal, on the other hand, there is in principle no end to the kind of evidence which might be relevant to the evaluation of a semantic theory. There are also all sorts of general constraints on good explanation which can be applied in semantics just as they can anywhere else.

I believe that there is much to be seen from the liberal perspective which might otherwise remain hidden. Since I see no way of saying, in advance of a search, what might count for or against a given semantic theory, I see no reason to think that such a search would be a waste of time. On the contrary, I think that a detailed examination of what may count for or against particular proposals can only bring illumination, both of the natural language itself and of the exigencies of understanding how that language works.

The liberal position will stand or fall by the quality and amount of the results yielded by research that follows its guidance. This paper as a whole may therefore be taken as a partial defense of liberalism.

I. FREGE'S PROBLEM

We turn now to semantics.

(1) Ralph believes that Tanya is a terrorist.
(2) Ralph believes that Patty is a terrorist.

Ralph has heard, from what he considers to be a reliable source, that Tanya is a terrorist. It so happens that Patty and Tanya are the same person. But poor Ralph does not know this, and is convinced that Patty is no terrorist. It would be wrong for me to report Ralph's belief that Tanya is a terrorist as the belief that Patty is a terrorist. Let us suppose that the wrongness of the report amounts to its falsity, rather than, say, its being merely misleading or lacking in explanatory value.

[2] The archetypical methodological conservative is Davidson. See in particular his "Belief and the Basis of Meaning," *Synthese*, XXVII (1974): 309–323. For a different version, see also Quine, "Methodological Reflections on Current Linguistic Theory," in Gilbert Harman and Davidson, eds., *The Semantics of Natural Language* (New York: Humanities, 1972).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The supposition is not mandatory. One could hold that, contrary to first appearances, intersubstitutions of coreferential expressions in the content sentences of propositional-attitude reports do preserve truth value. Appearances to the contrary would then be explained by pragmatic, rather than semantic, considerations. It would be worth exploring the consequences of such a view.[3] But that is not my present purpose. I take it that the view is prima facie less plausible than the alternatives, and it is these which I wish to explore in this paper.

So (1) and (2) may have different truth values, in spite of the fact that they are composed of expressions with the same referents in the same order. Frege's response to this was to claim that expressions embedded inside content sentences do not there refer to their normal referents, but rather to their senses. There are two claims here: first, expressions inside content sentences do not refer to their normal referents; and, second, expressions inside content sentences do refer to their normal senses.

If we follow Frege in supposing that embedded expressions do not refer to their normal referents, we are faced with a difficulty. Consider, if you would:

(3) Zippy believes that Zoe$_i$ is bright, and she$_i$ is bright.[4]

In (3), 'she' appears to be straightforwardly anaphoric with 'Zoe'. It looks as though the referent of 'she' is determined to be whatever its antecedent, 'Zoe', refers to. I have marked this with the subscripts 'i'. On Frege's theory, the referent of 'Zoe' in (3) is a sense. But the referent of 'she' ought not to be a sense, but the person, Zoe. We have a problem, however, with Zoe not being a sense.

Only three possible solutions come to mind: (A) appearances notwithstanding, 'Zoe' and 'she' both refer to a sense; (B) 'Zoe' refers to a sense, 'she' refers to Zoe; and (C) 'Zoe' and 'she' both refer to Zoe. (A) is certainly the least promising of the three options. If 'she' refers to a sense, then 'she is bright' should say that a sense is bright. We could make sense of this by adjusting the semantic value of 'is bright' to bring the truth conditions of the whole sentence into line. But this seems to be a singularly unmotivated move, given the availability of (B) and (C). And 'she' certainly does not behave as if it referred to a sense, since any expression that refers to Zoe may be substituted for

[3] For relevant discussion, see Nathan Salmon, *Frege's Puzzle* (Cambridge: MIT, 1986).

[4] 'She is bright' is meant to be outside the scope of 'believes'. In future examples, possible ambiguities are always to be resolved in that direction, so that the conjunction takes one of its conjuncts outside the scope of the doxastic operator.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

it without altering the truth value of 'she is bright'.[5] (B) and (C), on the other hand, are both prima facie quite plausible. I shall discuss each in turn with an eye to evaluating their prospects for incorporation into a satisfactory theory of meaning.

II. THE KAPLANESQUE SOLUTION

On a (B)-type solution to the problem, we would have to rewrite the anaphoric relation in (3). 'She' would not refer to the oblique referent of 'Zoe', but to something that stands in some special and specifiable relationship to that referent. How would this go? What we want to express is something like this. In (3), 'Zoe' refers to the sense of 'Zoe' and 'she' refers to the object that the sense of 'Zoe' presents. So we would need a general method of obtaining clauses of the form (4):

(4) R (e, b).

where e is an embedded expression and b is the object that the sense of e represents. The theory of meaning could then employ rules of the form (4) to predict that the referent of 'she' in (3) is whatever object stands in the relation R to the sense of 'Zoe'.

R is distantly related to David Kaplan's (and Alonzo Church's) relation Δ, so I call this version of a type-(B) solution the *Kaplanesque solution.*[6] But it is important to realize that Kaplan used his apparatus in a different way for a different purpose. The problem with the Kaplanesque solution is that it is not as easy to implement as one might think. Sometimes appropriate clauses of the form (4) are easy to come by:

(5) R ('Tanya', Tanya).

for example. But sometimes they are not. Consider:

(6) Dinsdale said that he loves [his wife$_i$], but he is rude to her$_i$.

How are we going to get a clause that allows us to recover the referent of 'her'? The simple disquotational axiom suggested by (5) will clearly not work here. What would we do with (7)?

(7) R ('his wife', his wife).

'His wife', unlike 'Tanya', does not have a fixed referent, but will have its referent determined by the surrounding linguistic context.

[5] Michael Dummett actually adopts a theory of this type. But in my view the semantics he works out is seriously defective. And he does little to make the general idea plausible. See his *Frege: Philosophy of Language* (London: Duckworth, 1973), ch. 9.

[6] See Kaplan, "Quantifying In," in Davidson and Jaakko Hintikka, eds., *Words and Objections* (Dordrecht: Reidel, 1969), pp. 178–214.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The theory of meaning thus cannot simply state which object is represented by the sense of 'his wife'. Rather, it will require some quite complicated rule that operates on 'his wife' together with other, nearby expressions to get to the right object.

And new complexities are presented by quantifier/pronoun constructions like those illustrated in (8).

(8)(i) Igor said he has [some yaks$_i$].
Olga will vaccinate them$_i$.
(ii) Zonker said he has [a new wife$_i$].
You will meet her$_i$.

Which relation holds between the sense of 'some yaks' in (8i), and the yaks that 'them' refers to? And which relation holds between the sense of 'a new wife' in (8ii) and the referent of 'her'?

I have no doubt that with ingenuity and hard work the dedicated Fregean could construct the appropriate mechanisms for associating semantic values with senses. But I think that, if we reflect upon precisely what information is required to get us to the right extensions, a non-Fregean line will appear more promising. In each of our examples, the appropriate extension for the anaphoric pronoun is recoverable from the semantic interpretation that the content sentence as a whole would have had, had it been treated as transparent and given its normal semantic interpretation. Thus, 'her' in (6) refers to what 'his wife' would have referred to, had (6) been referentially transparent.

Similarly in (8i-ii), the appropriate extensions for the pronouns would be recoverable from their antecedents, if we gave the entire content sentences their normal transparent interpretations. The theory of meaning will have the resources to recover the right extensions from transparent treatments of the content sentences, as is apparent from (9):[7]

(9)(i) Igor has [some yaks$_i$]. Olga will vaccinate them$_i$.
(ii) Zonker has [a new wife$_i$]. You will meet her$_i$.

Knowledge of what the interpretation of the content sentences would have been, had they been transparent, is thus sufficient to determine the extension of the anaphoric pronouns.

It is clear also that, in some cases, nothing less than that will do. In the examples (8), one has to know what the transparent interpretation of the entire content sentences would have been before one can figure out the referent of the unembedded pronouns. It is impossi-

[7] For an account of these resources, see Gareth Evans, "Pronouns, Quantifiers and Relative Clauses (1)," *Canadian Journal of Philosophy*, VII (September 1977): 467–563.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ble to figure out the extension of 'them' in (8i) without first ascertaining that 'them' refers to some yaks. And, to do this, one must give a transparent interpretation to the expression 'some yaks'. Moreover, one will only figure out that the particular yaks in question are the ones that Igor owns, if one interprets the two expressions 'he' and 'has' as if they had occurred transparently.

Similar remarks apply to (8ii). One will only be able to identify to whom 'her' refers, if one interprets each element in the content sentence as if it occurred transparently. To account for the anaphora in the sentences (8), what we need to do is thus precisely equivalent to treating the whole context transparently. Nothing more and nothing less than this is required. So, rather than attempt to formalize the Kaplanesque representation relation, R, I suggest that we look for some way of accounting for the failures of substitutivity while allowing the embedded expressions to keep their normal referents.

III. DAVIDSON'S ACCOUNT

If we adopt a (C)-type solution to our problem, and allow embedded expressions to keep their usual referents, we need to account for the failures of substitutivity. Davidson's account of the logical form of 'says that' offers a very attractive way of doing this.[8] Suppose that the logical form of 'Ralph said that Tanya is a terrorist' is really:

(10) Ralph said that. Tanya is a terrorist.

with 'that' being a demonstrative that refers to the utterance (the action of uttering) or inscription that follows it. The second sentence in (10) is treated as usual, its elements receiving their normal semantic values. The person who utters (10) is then making a claim about what Ralph said, a claim that can be understood roughly thus: '$(\exists x)$ (Ralph uttered x and, in virtue of x and my utterance of "Tanya is a terrorist," Ralph and I are samesayers').

If we assume that samesaying can appropriately relate utterers to believers, then we can apply Davidson's account to 'believes' sentences and explain the opacity of (1). (This assumption is not harmless; I shall return to the point later.)

'Tanya is a terrorist' is not very similar in saying power to 'Patty is a terrorist'. So Ralph may believe that:

Tanya is a terrorist.

but not believe that:

Patty is a terrorist.

[8] "On Saying That," *Synthese*, XIX (December 1968): 130–146.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Since each content sentence is also a sentence in its own right, we may assign to their component expressions their normal referents to solve our problems with coreference among expressions inside and outside content sentences.

Davidson's proposal is, as I said, attractive. It is attractive for two reasons. First, there is no invocation of senses; and the fewer invocations, the better, other things being equal. Second, it seems to read the semantics explicitly off the surface structure, or very nearly so. On Davidson's account, it is easy to see exactly what each familiar English word is doing and how each contributes to the whole. But only an incurable optimist could believe that such a neat and elegant theory could be true.

The proposal perhaps derives a modicum of prima facie plausibility from the fact that the indirect discourse 'that' and the demonstrative 'that' are spelled similarly. But this is quickly dispelled by noting that, in many languages (e.g., the Romance languages), the complementizer of indirect discourse and the demonstrative have completely different orthographic and phonetic forms. In some languages (e.g., Hebrew), the complementizer, unlike the demonstrative, does not even have the superficial form of a separate word, but rather of a mere affix.[9]

In English also, in spite of the orthographical coincidence, the demonstrative 'that' has a cluster of properties that the 'that' of 'believes that' does not. Notice first that they are pronounced differently. One usually says

(11) Ralph believes ðəţ Ortcutt is a spy.

shortening the 'that' to 'ðəţ'. The demonstrative 'that', however, seems never to undergo this phonetic reduction, or destressing. One cannot say *'Hey, look at ðəţ', or:

(12) *Ralph believes ðəţ. Ortcutt is a spy.

which interprets (11) in Davidson's way.

Second, the 'that' of 'believes that' and 'says that' is deletable. One can say:

(13) Lemmy said he needed money.

with the 'that' dropped. But the demonstrative 'that' cannot be dropped. You cannot say

(14) *Yes . . . is what I said.

[9] German and Finnish, on the other hand, are more like English in the relevant respect.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

 Digitized by Arya Samaj Foundation Chennai and eGangotri

with the 'that' deleted, instead of

(15) Yes that is what I said.

Now, there is not, to my knowledge, any well understood and generally accepted explanation of what determines the possibilities of deletion and destressing. But it is very plausible that, in both cases, the explanation will invoke the semantic role of the expressions involved. It looks as though the reason why one cannot delete the demonstrative pronoun 'that' is precisely because it has an important referring function. And similar remarks apply to the phonological characteristics. Stress often goes with semantic role, and it is hard to see what else could account for the stressing rules in this case.[10]

Although these objections, taken by themselves, are far from conclusive, they are very well worth noticing, for they provide a good example of how semantics can interact with other parts of linguistics. The 'that' of 'says that' can be deleted and destressed, but the English demonstrative pronoun cannot. This is simply empirical data. If it is true that the 'that' of 'says that' functions semantically like a demonstrative, as Davidson holds, then it must be false that what explains the phonological and syntactic facts is semantic role. Whether this is so or not is an empirical question at the intersection of the various parts of linguistics. The examples thus provide a nice demonstration of methodological liberalism at work, and point toward a rich field of evidence which may help us choose between competing semantic theories, should we have such aspirations.

There is another, more serious difficulty for the proposal that is purely semantic. The point of a theory of meaning is to explain why object-language sentences mean what they do. The explanation is effected by the assignment of logical forms and interpretations to the sentences, from which their meanings are deducible. A particular interpreted logical form entails a particular meaning. An assignment is correct only if the meaning of the sentence is the very meaning entailed by that assignment. And there are cases where it appears that Davidson's proposal fails in this crucial respect.[11]

(16) Everybody said that they loved Lucy.

is ambiguous. It can mean that everybody said of some group of

[10] For a discussion of these and other syntactic difficulties with Davidson's proposal, see M. Speas and Segal, "On Saying ðət," in *Mind and Language*, 1 (Summer 1986): 124–132.

[11] The objection that follows is due to James Higginbotham, though I put the point a little differently. See his "Linguistic Theory and Davidson's Program in Semantics," in Ernest LePore, ed., *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson* (New York: Blackwell, 1986), pp. 29–48.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

people that they loved Lucy. Each person would have said something like 'they love Lucy'. Or it can mean that each person said of himself (or herself) that he (or she) loved Lucy, saying something like: 'I love Lucy'. We might represent (16) in this second reading as

(17) Everybody$_i$ said that they$_i$ loved Lucy.

According to Davidson, at the level of logical form, (16) breaks into two quite distinct sentences. The first of the sentences contains a demonstrative 'that' that refers to what follows it. So the logical form of (16) should be the very same as that of (18):

(18) Everybody said that. They loved Lucy.

Indeed Davidson would say that the difference between (16) and (18) is merely orthographic, and 'without semantic significance' (p. 106). But (18) expresses only the first of the two possible meanings of (16). (18) simply cannot be used, by normal English speakers, to convey the meaning of (17). Imagine someone writing 'they loved Lucy' on a blackboard, pointing to it and saying 'everybody said that': how would you interpret this? So the difference between (16) and (18) is semantically significant.

Davidson's proposal could perhaps be adapted to cope with the ambiguity of (16) as follows. The ambiguity is the result of there being two different sorts of samesaying. When (16) reads like (17), the samesaying relation in question is one that holds between the utterer of 'they love Lucy' and utterers of 'I love Lucy'. But when (16) means the same as (18), the samesaying relation holds between the utterer of 'they love Lucy' and other utterers of 'they love Lucy'.

This suggestion is coherent. It is conceivable that English might work in the suggested fashion. The issue, however, is whether English actually does work like that. The fact that (18) is not understood by English speakers in the manner of (17) is evidence that it does not. If (16) had the same logical form as (18), then we ought to understand (18) exactly as we understand (16). This objection is not quite conclusive. Davidson could argue that, when (16) says the same as (17), its logical form, while still being paratactic, is not exactly the same as that of (18). And the difference between the two forms would account for the difference of meaning. The onus, however, is on Davidson to produce the required analysis. I shall try something else.

IV. HIGGINBOTHAM'S ACCOUNT

Higginbotham offers a proposal that is similar in spirit to Davidson's, but avoids some of its difficulties. The proposal is that we assign values to expressions only relative to the larger expressions of which

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

they form a part. So, if we have an expression e inside a larger expression, say, a sentence S, instead of assigning to e a value all by itself, we assign to e a value relative to S. The point of this is that, since an expression may simultaneously appear in several different embedding expressions, we may assign to it a value relative to one that differs from the value, if any, that we assign to it relative to another.

This tactic of assigning values only relative to embedding expressions is useful not just for attitude attributions, but can solve problems with sentences of a completely different kind.[12] This is important. If it were not so, then we would be introducing sweeping changes in the semantic theory, changes that affected what we said about every sentence, just to cope with the special problems with attitude attributions. That would seem like poor methodology.

The success of the proposal may ultimately depend upon just this. If the trick of making assignments of values relative to embedding expressions proves fruitful in a range of different cases, that will vindicate its use here. If, on the other hand, it idles and simply makes other areas needlessly cumbersome, then we should be suspicious.

The idea applies to belief attributions as follows. Consider (1) carved up as in (19):

(19) [S_1 Ralph believes [S' that [S_0 Tanya is a terrorist]]].

(19) is articulated into the three phrases, S_0, embedded in S', which is in turn embedded in S_1. In order to cope with cross reference, we want expressions inside content sentences to have their normal referents. So we treat the parts of the content sentence S_0, 'Tanya is a terrorist', relative to S_0 itself as if it were unembedded, and appeared in isolation. This means that the expressions inside the content sentence just get their normal referents as desired. This is expressed in (20). (I have altered Higginbotham's notation, for ease of assimilation.)

(20)(i) $\langle$'Tanya', $S_0\rangle$ refers to Tanya.[13]

(ii) $(\forall a)$ (a satisfies $\langle$'x is a terrorist', $S_0\rangle$ iff a is a terrorist).

from which it is easy to derive

(iii) $\langle S_0, S_0\rangle$ is true iff Tanya is a terrorist.

The angle brackets are to be taken literally, so that we are technically assigning values to ordered pairs of expressions. Intuitively we

[12] Specifically, certain sentence connectives, such as 'if' and 'unless', behave differently in different linguistic environments. See Higginbotham.

[13] My treatment of proper names here (and in example 5) is much too crude. But the crudity keeps things simple and is otherwise irrelevant to present affairs.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

can just understand ⟨'Tanya', S_0⟩ to mean ' "Tanya" considered relative to S_0'. It is easy to grasp the relationship between the predicates of the metalanguage which apply to the ordered pairs, and the similarly spelled English predicates 'refers to', 'is true', etc.

Now we look at S', 'that Tanya is a terrorist', and consider its role relative to the large sentence S_1, 'Ralph believes that Tanya is a terrorist'. S', considered relative to S_1, appears as the object of the verb 'believes'. Higginbotham (following Gilbert Harman[14]) takes the objects of propositional attitudes to be or to be objects "similar to" interpreted logical forms. Since the objects to which semantic values get assigned are themselves logical forms, the 'that' clause is taken to be true of things that are similar to itself. Similarity is used in conscious imitation of Davidson's samesaying relation: two things will be similar to each other if they have the same or similar semantic and pragmatic features. The role of the content sentence may be expressed by (20iv):

(iv) $(\forall a)$ (a satisfies ⟨S', S_1⟩ iff a is similar to S_0).

(iv) together with (v):

(v) ⟨'Ralph', S_1⟩ refers to Ralph.

and

(vi) $(\forall a)$ $(\forall b)$ (⟨a, b⟩ satisfies ⟨'x believes y', S_1⟩ iff a believes b).

along with simple interpretation rules give us (vii)

(vii) ⟨S_1, S_1⟩ is true iff $(\exists a)$ (a is similar to S_0 and Ralph believes a).

(vii) is a T-sentence that tells us, more or less, that 'Ralph believes that Tanya is a terrorist' is true iff Ralph believes something similar to 'Tanya is a terrorist'. Since this treatment has embedded expressions referring to their normal referents, we have a simple solution to the problem of anaphora at which we have been looking. Take (21):

(21) [S_1 Ralph believes that [S_0 Tanya$_i$ is a terrorist]] but [S_2 she$_i$ is too honest].

The challenge was to find some way of getting the unembedded 'she' in S_2 to refer to Tanya, so that S_2 could get its correct truth conditions. All we need to do is to find some suitable pair ⟨x, y⟩ such that:

(22) Ref ⟨x, y⟩ = Ref ⟨'she$_i$', S_2⟩ = Tanya.

[14] "Deep Structure as Logical Form," in Harman and Davidson, eds., pp. 25–47.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

But this is now no problem. 'Tanya' considered relative to the content sentence will just get its normal referent. So we have (23):

(23)(i) Ref $\langle$'Tanya$_i$', $S_0\rangle$ = Tanya.
(ii) Ref $\langle$'she$_i$', $S_2\rangle$ = Ref $\langle$'Tanya$_i$', $S_0\rangle$.

So,

(iii) Ref $\langle$'she$_i$', $S_2\rangle$ = Tanya.

The aspect of the theory which provides embedded expressions with their normal referents seems to be satisfactory. The other aspect of the theory, however, which is supposed to account for the failures of substitutivity, is less so. How are these failures to be explained?

'Ralph believes that Tanya is a terrorist' is true, but 'Ralph believes that Patty is a terrorist' is false. Higginbotham's account allows for this easily enough. The content sentence, considered relative to its embedding-attitude report, is treated as a predicate true of things similar to itself. The content sentences 'Tanya is a terrorist' and 'Patty is a terrorist' are not similar to each other and obviously might fail to be similar to the same things. Since each is true of things similar to itself, they will be true of different things. One may be similar to, and so true of, what Ralph believes, whereas the other is not.

Similarity appears here as a technical notion within the theory of meaning: it is what determines the extensions of the content sentences in their predicational roles. For this reason, we need to be clear about exactly what similarity is doing and how it is supposed to be doing it.

At this juncture it will be well to reflect upon how samesaying works for Davidson.[15] If I say that Galileo said that the earth moves, then samesaying relates my utterance of the content sentence to a similar sort of thing. (Strictly speaking, samesaying relates people, not utterances; but I think the extrapolation is safe.) To arrive at something that samesays with Galileo's utterance, proceed as follows. Take Galileo's utterance. Discern within it an articulated linguistic structure of Italian. Apply to it a truth theory (for Italian, in English). This theory will deliver English sentences that have the same truth conditions as the Italian one. Now, there will be various different ways of stating the conditions under which Galileo's utterance is true, differences which will be too subtle for the truth theory to note. But among these options will be some that are better suited to capturing the flavor of Galileo's utterance—more seventeenth-century

[15] See, e.g., his 1976 paper, "Reply to Foster," reprinted in *Inquiries into Truth and Interpretation* (New York: Oxford, 1984), pp. 171–179.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Italy, less twentieth-century America. Those suitably flavored ways of stating the truth conditions of Galileo's utterance samesay with it.

Davidson's samesaying does not appear within the theory of meaning. But it helps us to understand how, for example, I can say what Galileo said by uttering an utterance that differs from Galileo's. In uttering his utterance Galileo said something, and there are various different ways in which I can report what he said. I can pick any utterance, *u*, that samesays with Galileo's, and use *u* to say what Galileo said.

There is a superficially analogous phenomenon with belief reporting. Ralph has the particular belief that Tanya is a terrorist. And there are various ways in which I can say what Ralph believes. I can pick any utterance, *u*, that has the same content as Ralph's belief, and use *u* to say what Ralph believes. It is this phenomenon, I take it, which Higginbotham's similarity relation is supposed to capture. Similarity helps to explain why there is more than one way of saying what Ralph believes. The idea is that various sentences bear a similarity relation to what Ralph believes and any of these may appear as the content sentence of a true belief report.

The idea is reflected in the theory of meaning by the *T*-sentence for (1):

(20-vii) $\langle S_1, S_1 \rangle$ is true iff $(\exists a)$ (a is similar to S_0 and Ralph believes a).

'a' ranges over interpreted logical forms and 'S_0' refers to the logical form of 'Tanya is a terrorist' (call this logical form 'LF(TT)').[16] So, according to (20-vii), there is some quite specific logical form which Ralph actually believes and which is similar to LF(TT). But is this right? Of all the logical forms that are similar to LF(TT), is there really one special one that he believes? If so, which one is it?

The difficulty here is that believing is not like speaking. When Ralph utters an utterance, he produces a specific linguistic object. This utterance can be plugged into a theory of meaning, and the theory will then tell us what its logical form is. But you cannot plug a belief into a theory of meaning to find out what its logical form is. To associate a logical form with a belief you must first find a sentence that gives the content of the belief. The theory of meaning can then work on the sentence to produce a logical form. But there are many sentences that correctly give the content of Ralph's belief: the con-

[16] Actually, Higginbotham says that the objects of belief are interpreted logical forms or things similar to these. But, since I do not know what sorts of things other than logical forms are similar to logical forms, I shall just suppose that 'a' ranges over logical forms.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tent sentences of all true reports of that belief. So Ralph believes the logical form of the content sentence of any true report of his belief.

The disanalogy with 'says that' on Davidson's construal is plain. If 'Ralph says that Tanya is a terrorist' is true, then Ralph utters a single, specific utterance that samesays with an utterance of 'Tanya is a terrorist'. But uttering such an utterance is one way of saying that Tanya is a terrorist. Ralph does not merely say something that samesays with 'Tanya is a terrorist', he actually says that Tanya is a terrorist, does he not? If Higginbotham's treatment of 'believes' is to work in a similar fashion, the analogue should be this: if 'Ralph believes that Tanya is a terrorist' is true, then Ralph believes* some single, specific logical form that samesays with LF(TT). Believing* a logical form would stand to believing something as uttering an utterance stands to saying something. And believing* a logical form that samesays with LF(TT) would be just one way of believing that Tanya is a terrorist.

The *T*-sentence for (1) ought, then, to read: (1) is true iff Ralph believes* some logical from similar to LF(TT). But, if this is to mean anything, we need an account of believes*. And I, at least, have no idea what this account would be.[17] I just do not know what believing* would be. If the semantics is to get by with belief instead of belief*, similarity adds an unnecessary wrinkle to the theory. Since Ralph must believe the logical form of any content sentence of a true belief report, Ralph does not only believe something similar to LF(TT), he also believes LF(TT) itself. Higginbotham's *T*-sentence for (1) is still true; after all, Ralph does believe something similar to LF(TT). But this is also true: (1) is true iff Ralph believes LF(TT). And this, I claim, provides a much tidier description of the role of the content sentence (considered relative to the embedding report) than Higginbotham's original (20vii).

V. FREGE VINDICATED?

Since (1) is true iff Ralph believes LF(TT) itself, there should be no objection to viewing content sentences as singular terms that refer to their own interpreted logical forms. I suggest we treat the logical forms of content sentences as complex self-referential devices. Logical forms are structured items, and each logical form is structurally isomorphic with itself. This opens up the possibility of a composi-

[17] Perhaps to believe* something is to have a sentence of brain language playing a particular computational role in your mind/brain. But, if this is so, it is an empirical truth that has nothing to do with the semantics of 'believes'. A semantics for belief attributions should work in a world where we did not have sentences in our heads. 'Believes*', remember, would be a technical term in the theory of meaning, not the theory of belief.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tional account of the semantics of content sentences (considered relative to their embedding reports). Each simple element of a logical form will refer to itself. It should then be possible to predict the semantic value of a complex logical form, including that of a whole content sentence, from its structure and the values of its parts.

So, instead of adopting (20iv), I think we should attempt to construct a theory along these (roughly sketched) lines:

(24)(i) ⟨'Tanya', S_1⟩ refers to *Tanya*.
(ii) ⟨'is a terrorist', S_1⟩ refers to *is a terrorist*.

I assume that S' just inherits the value of S_0 considered relative to S_1. This should allow us to derive (iii):

(iii) ⟨S', S_1⟩ refers to *Tanya^is a terrorist*.

The italics signal that the expressions refer to components of interpreted logical forms, rather than their normal referents; and '^' signifies the correlate of concatenation at the level of logical form. The elements of the form go together in the way required by the theory of meaning.

(24iii) together with the original (20v and 20vi) suffice to yield (25):

(25) ⟨S_1, S_1⟩ is true iff Ralph believes *Tanya^is a terrorist*.

That is to say, 'Ralph believes that Tanya is a terrorist' is true iff Ralph believes the interpreted logical form of the sentence 'Tanya is a terrorist'.

This may seem a little odd. What does it mean to say that believers believe interpreted logical forms? And, you might wonder, how could Galileo believe the interpreted logical form of an English sentence that he has never heard? My answers are implicit in what I have already said. Galileo believes the interpreted logical form of 'the Earth moves' because 'the Earth moves', with its proper English interpretation, correctly gives the content of his belief. In general, a subject believes the interpreted logical forms of all and only those sentences which correctly give the contents of her beliefs. It is not up to the theorist of meaning to say which sentences these are: that is a question for psychologists. It is just the question of how best to describe someone's beliefs.

My proposal has three obvious advantages over Higginbotham's, one minor, two major. First (the minor one), on my account, the 'that' clauses are treated as what they appear to be, complex singular terms that refer to the objects of belief. They are not treated as quantifier/variable constructions, which they do not appear to be.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Second, my proposal promises a compositional account of the semantics of the content sentences relative to their surrounding environments. On Higginbotham's view, the content sentences appeared as semantic monoliths, and their internal structure was simply ignored. Third, my proposal has no use for similarity. We have already seen that, if similarity is to do any work for Higginbotham, then it must be employed together with an account of belief*. Lacking such an account, similarity adds nothing to the theory but unnecessary complications. With (25) serving as the *T*-sentence for (1), in place of (20vii), similarity is exorcised.

Similarity is fecund, but it is also notoriously vague and unhelpful. I think that similarity does have a role in the explanation of our intuitions about permissible intersubstitutions inside content sentences. But I think that these intuitions have more to do with our views about the nature of belief than they do with our grasp of language. So similarity ought to be kept out of the theory of meaning, which is concerned with precisely that. Let me explain.

The logical forms of two orthographically distinct sentences are never the same. For the logical form of a sentence includes the words in it [see, for example, (19), which gives a partial description of the logical form of (1)]. This means that different expressions are never safely intersubstitutable inside embedded sentences, even if they are synonymous. Since what the embedded words refer to includes the words themselves, different embedded words will always have different referents. So, for example,

(26) Horace believes that a fortnight is a fortnight.

will not entail

(27) Horace believes that a fortnight is two weeks.

I think that this is a desirable consequence of my account. Intuitions vary about whether (27) does follow from (26). There are those who would claim that it does not.[18] These people are not lunatics: that is to say, they are not denying a logical truth. Yet there are those, and I am one, who think that (27) does follow from (26). And I am not a lunatic either.

What I believe, more precisely, is that (27) follows from (26) along with (28) and (29):

(28) $(\forall p)(\forall q)(\forall x)((x \text{ believes } p \ \& \ p \text{ is similar to } q) \rightarrow x \text{ believes } q)$.

('x' ranges over believers, 'p' and 'q' range over interpreted logical forms).

[18] Cf. Tyler Burge, "Belief and Synonymy," this JOURNAL, LXXV, 3 (March 1978): 119–138.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(29) The logical form of 'a fortnight is a fortnight' is similar to that of 'a fortnight is two weeks'.

Given (28) and (29), one may infer (27) from (26). Those who reject the entailment do so because they reject the principle stated in (28), or because they do not believe (29).

(28) will license intersubstitutions of those embedded expressions which are appropriately similar in content. It is up to those of us who accept (28) to determine what counts as similarity. 'Similarity' is used here merely as part of an analysis of why some people are inclined to think that, for example, if (26) is true, then so is (27). Since (26) does not by itself entail (27), this has nothing directly to do with the theory of meaning.

I conclude now with some remarks about Frege. It should be clear that my account is Fregean in spirit. Interpreted logical forms may with some justice be called 'senses'. They do not have all the features that Frege associated with senses, but they have two: they are the objects of propositional attitudes, and they serve as oblique referents for embedded expressions. This second function gives us a completely Fregean account of the failures of substitutivity.

Higginbotham's tactic of relativizing assignments of semantic values to linguistic contexts allows us to solve the problem of cross reference across the boundaries of opaque contexts. But even here we have not progressed much beyond Frege, who said in another (but not unrelated) context: "This shows that the subordinate clause in our original complex sentence is to be taken twice over, with different referents: once for a thought, once for a truth value."[19]

GABRIEL SEGAL

University of Wisconsin/Madison

[19] "On Sense and Reference," Max Black, trans., in Peter Geach and Max Black, eds., *Translations from the Philosophical Writings of Gottlob Frege* (New York: Blackwell, 1952), p. 76.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

A PARADOXICAL DEPARTURE FROM CONSEQUENTIALISM*

IN apparent agreement with Bernard Williams that consequentialism faces insuperable difficulties with the issue of personal integrity, Samuel Scheffler[1] has argued for what he calls an agent–centered prerogative. The prerogative differs from consequentialism in that it allows as permissible what consequentialism holds is always wrong, namely, that an agent may sometimes act in a way that is *not* designed to bring about the best available state of affairs on impersonal moral criteria. But, since the prerogative retains the consequentialist feature that it is always permissible to act on fully consequentialist criteria, Scheffler's view entails the rejection of agent-centered restrictions that make it sometimes wrong to act in fully consequentialist ways.

I argue here that Scheffler's use of the notion of integrity as the rationale for the prerogative departure from consequentialism is paradoxical on any plausible interpretation of the prerogative's application. I begin with a formulation of Scheffler's prerogative, and then try to show the paradox in accepting it and rejecting two agent-centered restrictions involving personal promises and personal relations.

I. THE PREROGATIVE AND ITS RATIONALE

Scheffler's prerogative assumes a set of impersonal moral criteria for ranking states of affairs from best to worst. Any conceptual scheme that requires that it is at least sometimes obligatory to bring about the best available, impersonally specifiable state of affairs requires such a set of criteria. And it does not matter here whether the criteria are, in Ronald Dworkin's terms, goal-based, as in the principle of utility, or rights-based, as in a Rawlsian theory. All that is required is that the criteria provide an impersonal basis for ranking states of affairs from best to worst.

If we call the *impersonal criteria* the IC, we can state Scheffler's prerogative in the following way: when an agent is faced with a decision between several courses of action with various outcomes that can be ranked from best to worst on the IC, the agent may

* I am indebted to Paul Hughes, the late Mark Overvold, and Peter Smith for helpful comments on earlier versions of this paper. I also wish to thank my wife, Patty Harris, for tirelessly following every revision—no matter how small—that has been made in this paper since its inception.

[1] *The Rejection of Consequentialism: A Philosophical Investigation of the Considerations Underlying Rival Moral Conceptions* (New York: Oxford, 1982).

0022-362X/89/8602/90–102
© 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

assign—within some designated proportion—a comparatively greater weight to his own personal projects in measuring outcomes than that afforded to impersonal outcomes. He may then choose that action which has a nonoptimal impersonal outcome and the optimal outcome for his own personal projects as long as the impersonal costs do not exceed the designated proportion. Only if all the alternative actions available to the agent have outcomes for his personal projects which exceed the designated proportion is he required to act to bring about the state of affairs ranked highest on the IC. Otherwise, it is permissible within the parameters of the prerogative to depart from the IC and fully consequentialist considerations. And, finally, since it is a prerogative, it is *always* permissible and *never* wrong for an agent to abandon his personal projects to act solely on the fully consequentialist considerations of the IC.

Scheffler's prerogative departure from consequentialism is designed to allow a moral agent—within restrictions—to devote time and energies away from the impersonal project of fully consequentialist morality and the IC to the ground or basic projects that give unity and personal meaning to his life. This is not to deny that devotion to the IC can be a basic personal project. But this is the exception. The typical person lacking access to any possible basic project except devotion to the IC would probably lack reasons for living. And the person who is forced to be attached to the IC through stringent expectations, associated guilt, shame, etc., will be alienated from his life and lack a sense of integrity which has personal content.

Scheffler's prerogative, on the other hand, allows the agent to recognize that impersonal demands can be placed on his life, but that his life has an independent personal grounding of its own. Moreover, it is always permissible for any person to attach himself personally to the IC as his own basic project, or to depart from his personal projects on any occasion to act in accordance with the IC. The aim, of course, is to replace what is perceived as an overly demanding and alienating conception of morality with a conception which is demanding but which allows for a personally integrated life.

II. THE PARADOX

Scheffler claims that there are two problems with agent-centered restrictions: they have an air of paradox about them and they lack a rationale. While undoubtedly true of some alleged agent–centered restrictions, it is a mistake to treat all such restrictions as cut from the same cloth. I shall not directly address the first charge here. Rather, I shall formulate two restrictions—the personal-promise restriction and the personal-relations restriction—and argue that, if Scheffler's prerogative departure from consequentialism has a rationale (that it

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

accommodates important features of personal integrity), then these two restrictions also have a rationale, since they are necessary to insure forms of personal integrity allowable under any plausible interpretation of the prerogative and the conditions of its application.

The two restrictions. On any plausible interpretation of the prerogative and the conditions of its application, there will be allowable personal projects that involve personal *commitments*—commitments that do not have their origin in IC considerations. Some of these commitments will involve personal promises and others will involve other forms of interpersonal relations. The point here is that the forms of personal integrity which make life worth living for most people involve personal projects that are *interpersonally dependent.* They are not socially isolated projects that can be carried out in a vacuum.

Now let us assume, with Scheffler, that any personal project that is allowable under the conditions of the prerogative must come at no greater cost than the designated proportion. Call the range of this cost the P-*range cost* and any personal project that falls within it a P-*range project.* If what I have said so far is correct, then any plausible interpretation of the prerogative and the conditions of its application will allow that there are some interpersonally dependent projects that involve personal commitments that are P-range projects.

The personal-promise restriction asserts (a) that a person can only make personal promises, the keeping of which comes within a P-range cost, and (b) that, in anything like normal circumstances, he must keep all promises that are made within P-range projects, especially promises, the breaking of which would undermine the dependent P-range projects of the person to whom they were made. The personal-promise restriction is an agent-centered restriction, because it says that the violation of some personal promises is wrong even if violating the promises would bring about the state of affairs ranked highest on the IC.

The personal-relations restriction asserts (a) that an agent may voluntarily enter a personal relationship with someone only if the relationship comes within a P-range cost, and (b) that, in anything like normal circumstances, he must attend to any personal relationship that is such that, (i) if it were ignored or neglected, the relationship would be undermined, (ii) it is someone's relationship to the agent that is a part of a P-range project for that person as well as the agent, and (iii) it is a relationship that the agent has voluntarily entered. Once again, the personal–relations restriction is an agent-centered restriction, because it says that the abandonment of some

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

personal relations is wrong even if abandoning those relationships would bring about the state of affairs ranked highest on the IC.

Enough has been said to establish the point that, on any plausible interpretation of the application of the prerogative, the making and honoring of some personal promises will be *permissible* where such promises on the part of person *A* are vital to the basic projects of person *B* and where the basic projects of person *B* are P-range projects. The problem for Scheffler is to admit that promises made within P-range projects are permissible, but honoring them is not *obligatory*. Herein lies the paradox of Scheffler's departure from consequentialism.

Call any promise that falls within a P-range project allowable under Scheffler's prerogative a P-*range promise* and any project dependent upon such a promise a P-*range promissory project,* and assume the following:

(1) P-range promissory projects are morally permissible.
(2) *B* seeks a promise from *A*, the keeping of which is necessary to *B*'s basic project.
(3) *B*'s basic project is a P-range promissory project relative to the IC—it is morally permissible.
(4) *A*'s making the promise to *B* is permissible but not obligatory—it is a P-range promise.
(5) If *A* makes the promise to *B* and the promise can be kept under normal moral circumstances, then *A* puts himself under the obligation to keep the promise to *B* under normal moral circumstances.

(The concept of "normal moral circumstances" and its interpretation are crucial to the issue of whether or not this premise and premise (6) is acceptable. I shall address these issues after completing the current argument.)

(6) The keeping of *A*'s P-range promise to *B* can be done under normal moral circumstances, that is, it is not an abnormal moral circumstance that the keeping of a promise results in a less than impersonally optimal outcome as long as the cost of its doing so is a P-range cost.[2]

[2] Scheffler cannot say that keeping a P-range promise is done at an abnormal moral cost, for his prerogative is designed to allow as a part of the normal moral life projects that are not strictly proportional to the requirements of the IC. If keeping the promise is *permissible* under normal moral circumstances at a P-range cost, then the IC *cost* of keeping the promise is not a feature under P-range conditions for claiming a change in normal moral circumstances as a ground for not keeping the promise.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

[As with premise (5), I shall defend (6) against objections after completing the current argument.]

(7) Therefore, if *B*'s P-range promissory project is morally permissible, then, if *A* makes the P-range promise upon which it depends, then *A*'s keeping the promise is morally obligatory.
(8) *A* makes the promise to *B*.
(9) *A* correctly believes that his breaking the promise to *B* would result in a state of affairs ranked highest on the IC.

From (9) and Scheffler's prerogative, it follows that

(10) It is not obligatory for *A* to keep the promise to *B*, even though keeping the promise comes within a P-range cost.

And from (10) and (7) it follows that

(11) *B*'s P-range promissory project is not morally permissible.

which contradicts (3), which says that it is.

If we reasonably assume that there is always (or at least generally) more that could be done in the name of any plausible interpretation of the IC, we can generalize to the conclusion that

(12) P-range promissory projects are not morally permissible.

This, of course, contradicts (1), which is an assumption of any reasonable application of Scheffler's prerogative.

Objections might arise, however, concerning the use of the concept of "normal moral circumstances" and how that concept figures into the acceptability of premises (5) and (6). Some clarification, then, is in order.

Part of what it is for *A* to put himself under an obligation to *B* by making a promise to *B* includes at least an implicit understanding between them regarding the limits and scope of *A*'s incurred obligations in relationship to *B*'s expectations as made legitimate by the former's commitment. To understand these conditions is to understand the "normal moral circumstances" under which the fact that *A* has made a promise to *B* results in his actually having the obligation to keep it. This seems entirely uncontroversial to me.

Interpretations will differ, however, on what counts as normal moral circumstances by differing on what counts as a morally permissible reason for not keeping a promise that has been made. For example, I do not think that any reasonable person would deny that any acceptable conception of promises includes the understanding that some promises can be broken to avoid certain catastrophes. Given this understanding of permissible reasons for breaking a

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

promise, it is a change in the normal moral circumstances that an unforeseen catastrophe could be averted by not keeping the promise. This change in the normal moral circumstances releases the promisor from the obligation to keep the promise, and this is an implicit understanding of any reasonable conception of the obligations incurred by making promises. But there is certainly no ground for disagreement between Scheffler and the personal-promise restriction on this score.

The issue, rather, is this: How might there be other competing interpretations of the concept of normal moral circumstances and morally justifiable reasons for breaking promises such that, on one view, premises (5) and (6) are true and, on another—namely, Scheffler's—at least one of these premises is false. And, moreover, what is the argument for the interpretation that makes (5) and (6) true?

Scheffler does suggest an interpretation of the prerogative which would make it wrong to break P-range promises unless doing so would result in a better state of affairs.[3] Ruled out on this view is an agent's breaking a personal promise merely to pursue some other personal project of his own. Still allowed are promises broken to advance states of affairs ranked higher on the IC than those advanced by keeping the promise.

On this view, breaking a P-range promise for personal reasons is not a morally justifiable reason, and a change in personal circumstances in this sense does not translate into a change in normal moral circumstances. But the discovery, after having made a P-range promise, that there is an impersonally superior alternative to keeping the promise *does* count as a change in normal moral circumstances. Moreover, a mere change in mind where the promisor already knew of the impersonally superior alternative *before* making the promise would also count as such a change. If this interpretation of "a change in normal moral circumstances" is acceptable, then premise (6) is false in the event that *A* changes his mind about the promise, not in favor of his own personal projects but in favor of IC considerations.

But there is something very troubling about this interpretation of the circumstances assumed to be normal for an acceptable conception of promises and their incurred obligations. The troublesome thing about it is not just that it jars with our intuitions. It is deeper than that. Scheffler's views could be counterintuitive without being paradoxical, and I have made the latter, stronger claim. What I have claimed is that, on any plausible interpretation of what the preroga-

[3] *The Rejection of Consequentialism,* fn. 2, p. 85.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tive would allow, his prerogative departure from consequentialism is paradoxical. Therefore, I must argue for an interpretation of what counts as normal moral circumstances and morally permissible reasons for breaking a promise which is more stringent than Scheffler's interpretation. And I must argue that any plausible construal of the prerogative would allow for commitments that reflect this more stringent interpretation, one that makes (5) and (6) true.

Scheffler's restriction on breaking promises is too limited. It takes the following form. It is morally wrong for *A* to break a P-range promise by giving priority to *A*'s own interests over those of *B*, contrary to the indications of the promise. On this view, *A*'s making a promise to *B* includes only a subordination of his own interests to those of *B*. Surely this is what it means to make a promise in some moral circumstances and *part* of what it means in others. But many typical promises that *A* might make to *B* would include far more than this. They would include a commitment by *A* not only to the priority of the interests of *B* over the interests of *A*, but to *B*'s interests over the interests of some other person(s), *C*, let us say, as well. Thus, a more stringent interpretation of such commitments which would make premises (5) and (6) true is one in which the condition of *C* is *to some degree* worse off on relevant IC considerations than that of *B*, and yet the commitment is to *B*.

I do not see any reason to believe that these latter kinds of promises are any less prevalent or important than the former. After all, it is only one kind of moral complaint about broken promises that the promisor cares more about *himself* than his commitments. Surely, a prior commitment in the form of a promise precludes the permissibility of making a similar commitment to another person when both commitments cannot be honored. The fact, then, that *A* has subsequently (and forgetfully?) made a promise to *C* similar to the one made to *B* cannot plausibly count as a relevant change in normal moral circumstances—a change that would release *A* of the obligation to *B*—despite the fact that *A* would not be breaking the promise to *B* to pursue his own interests. A conception of promises which *did* allow this as a relevant change in moral circumstances would not only be counterintuitive, it would be a view of such commitments which is too shallow to allow adequately for the kinds of raised expectations and interpersonal dependencies which we think that promises should accommodate.

Just as implausible is a conception of promises which releases one from the obligation to keep a P-range promise the moment the promisor comes upon another person minimally worse off than the promisee on IC considerations. It is implausible, not merely because it is counterintuitive *vis-á-vis* our prevailing conception of prom-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ises, but because the concept of commitment it embodies is too shallow to allow for the kinds of raised expectations and interpersonal dependencies which should plausibly be made available by the prerogative.

Let me illustrate with an example, one that does not necessarily involve a basic project but does involve important considerations. Sue is driving on a moderately crowded highway, when she comes upon a stranger with a flat tire in need of help. She is in somewhat of a hurry because she has some P-range plans for the evening exclusively involving her own interests. She decides, however, that she has just enough time to help and decides to do so, knowing that in this city other persons are not likely to render aid. She also knows that there are often motorists in need of similar assistance along this road, more than one could ever help. The stranded motorist does not have a spare, so someone must take the flat tire to a service station to have it repaired and someone must stay with the car to prevent vandalism. Since the motorist does not know how to drive Sue's car because of its transmission, she promises to take the tire to the station and return with it. The motorist agrees, and Sue proceeds with the tire to the service station.

Along the way, Sue comes upon a second stranded motorist with *two* flat tires (or something of the sort) but otherwise relevantly similar to the first motorist. She becomes frustrated. She correctly believes that she can now do only one of three things: (i) she can keep the promise to the first motorist, (ii) she can break the promise and proceed on to her P-range plans for the evening, or (iii) she can break the promise and help the second motorist. Let us assume that the moral disappointment of the first motorist over the broken promise would not be any greater than the moral disappointment of the second motorist over the perceived indifference of Sue's not stopping to render aid, and that the loss of the first motorist's tire would not be quite as great as having two flat tires, or whatever.

The concept of commitment embodied in Scheffler's concept of what constitutes a change in normal moral circumstances rules out option (ii), but it does nothing in the way of making sense of Sue's having made a commitment to the first motorist which takes moral priority over the interests of the second motorist. And, on Scheffler's view, it would be wrong for the first motorist to harbor resentment toward or lodge moral complaints against Sue if she decided to help the second motorist. Moreover, under these circumstances, what appeared to be P-range expectations of the first motorist are not permissible if Sue decides to help the second motorist.

This notion of promises and what counts as a change in normal moral circumstances is not only counterintuitive, it is simply too

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

shallow in terms of a concept of commitment to make sense of the legitimate expectations and the P-range promissory projects of the promisee. This is especially true in a world rife with both the room and opportunity for improvement on IC grounds and where some of the dependent projects of the promisee are central to his integrity.

Paradoxically, then, Scheffler's prerogative leads to the conclusion that a person can make a P-range promise upon which the P-range basic projects and legitimate personal integrity of another person depend, yet have no obligation to keep it. The only exception is where there is no impersonally superior alternative to keeping the promise. Thus, avoiding the paradox in a way that preserves the permissibility of projects such as those of *B* requires denying the permissibility of *A*'s ignoring such voluntarily made P-range promises to further impersonally specified ends. This is what the personal-promise restriction does, which entails that it would sometimes be wrong for *A* to act on impersonal optimizing criteria. This is an agent-centered restriction with the same rationale that motivated the prerogative departure from consequentialism in the first place. Thus, it is paradoxical to accept personal integrity as the rationale for the prerogative and deny it for the personal-promise restriction.

Now, consider the analysis of personal relations. On any plausible interpretation of the application of the prerogative, persons will be allowed to have special relations with other specific individuals as important aspects of their basic projects. This is because for most people there is no basic project which excludes the relations of friends, lovers, and family which would make life worth living for them. Call any personal relationship an *R-relationship,* and any R-relationship that is part of a P-range project a *P-range relationship.*

Before constructing the argument about P-range R-relationships similar to the one about P-range promises, let me say something about the relationships among the concepts of R-relationships, commitments, and obligations. Not all special R-relationships involve the notion of commitment. Two people, for example, might spend a great deal of time together as drinking partners without any degree of special commitment to each other, simply because they enjoy each other's company. To say otherwise is to trivialize the concept of commitment. Other R-relationships involve commitments on the part of one or all the parties to the relationship. The devoted but unrequited lover and the anonymous benefactor represent persons who are committed to other special persons without a mutual commitment. Yet other personal relationships involve mutual commitments. Marriage and friendship are two good examples.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

It is not at all clear that all R-relationships involve special obligations. Nor is it clear that all R-relationships that involve commitment involve special obligations. For example, the commitment of the devoted but unrequited lover makes perfect sense apart from any concept of his obligations to his lover. The same can be said for the unknown benefactor or for those who have a supererogatory dedication to particular persons. There are, however, some R-relationships that involve commitments that do not make sense apart from the concept of incurred obligations, and they do not all involve mutual commitments. What is it, for example, to decide to become a parent in more than a biological sense, apart from the idea of incurring parental obligations that carry a great deal of weight in deciding what one must do? And, in the case of young children, this relationship can hardly be thought of as a mutual commitment.

I choose here, however, to focus on R-relationships that involve mutual commitments. Perhaps there are mutual commitments that make sense apart from the notion of incurred obligations, though I doubt it. What is undeniable is that there *are* mutual commitments that do *not* make sense apart from the notion of incurred obligation. Not all marriages involve the concept of a contract, especially those made in heaven, but even those made in heaven involve the notions of mutual commitment and incurred obligations, especially those lived on earth. And, if one can free oneself from a certain legalistic way of thinking about morality and commitment, it is easy enough to see that the same normative features apply to friendship, family, and other similar relations. In the case of these R-relationships, mutual commitment without mutual obligation is simply an incoherent concept.

With these kinds of R-relationships in mind, consider the following argument regarding Scheffler's prerogative as it applies to two agents, A and B.

(1) P-range R-relationships are morally permissible.
(2) Entering R is not obligatory for either A or B.
(3) R is permissible for A and B if R is a P-range relationship (where entering R and sustaining R come within a P-range cost) and A and B voluntarily enter R with each other.[4]
(4) R is a P-range relationship for A and B.
(5) A and B voluntarily enter R with each other.
(6) R, as a voluntarily entered P-range relationship, is permissible for A and B.

[4] I am not asserting here that all mutual R-relationships must be voluntarily entered to be permissible or to be P-range relationships. Rather, I am asserting that some of them must and that the present argument is meant to apply to them.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(7) If R is a permissible P-range relationship for A and B, then, by voluntarily entering R, A incurs the obligation to B under normal moral circumstances to do the things that sustain R with B, and B incurs the obligation to A under normal moral circumstances to do the things that sustain R with A.

Nothing is said here about what counts as normal moral circumstances for different R-relationships that involve mutual commitments. If the R-relationship between A and B is based on a mutual love, for example, it might be claimed with some plausibility that, if A ceases to love B, this counts as a change in the circumstances such that it no longer makes sense to speak of the relationship and commitment obtaining under normal moral circumstances. But even here the commitment does not disappear the instant the feelings are gone. Any plausible view of what a mutual commitment is contains some considerations *as a part of the commitment* as to how the parties should treat each other should the relationship end. These considerations will be among the obligations incurred by the commitment. That they are not clearly set out in advance and that it is difficult to specify what these obligations are for various R-relationships and for various individuals involved in such relationships should not detract at all from the conviction that they obtain.

(8) Under normal moral circumstances, doing the things that will sustain the R-relationship between A and B will not bring about the state of affairs ranked highest on the IC but will come within a P-range cost.

(9) A correctly believes that not doing the things that sustain R with B will bring about the available state of affairs ranked highest on the IC.

From (9) and Scheffler's prerogative, it follows that

(10) A does not have the obligation to do the things that sustain R with B.

From (10) and (7) it follows that (11) the voluntarily entered P-range R-relationship between A and B is not permissible, which contradicts (6), which says that it is permissible.

Once again, if we reasonably assume that there is always more that could be done in the name of any plausible interpretation of the IC, we can generalize to the conclusion that

(12) P-range R-relationships (at least of the sort in question) are not morally permissible.

This contradicts (1), an assumption of any reasonable application of Scheffler's prerogative.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The only way to avoid the paradox *and* secure the permissibility of R-relationships between *A* and *B* as basic projects is to deny the permissibility of *A*'s ignoring the P-range R-relationships he has voluntarily entered under the conditions of the prerogative. And the restriction must include a prohibition on *A*'s ignoring his P-range R-relationship with *B* to further impersonally specified ends. This is what the personal-relations restriction does, which, again, entails that it would sometimes be wrong for *A* to act on impersonal optimizing criteria.

It might be objected, however, that the concept of assumed moral risk can resolve the apparent paradox. The idea would be that a person would be morally justified in taking on such interpersonally dependent projects, if he were willing to assume the risks to his own personal integrity should the project fail because the other person decided to abandon the project. In this way, embarking upon these projects would be permissible and consistent with both persons accepting Scheffler's prerogative.

Three things must be said here. First, the notion of assumed risk in this context is not the notion that *B* must assume the risk of *A* not doing what *A* ought to do, but rather of *A* not doing what *A* may permissibly do. Second, given what is at risk, namely, forms of personal integrity central to everyday life, this notion of assumed moral risk is just an implausible moral concept. For, in the case of a family whose lives were shattered by their being abandoned (within a P-range cost) in the name of utility or some other IC consideration, it would lead us to say that their plight was their own responsibility. This is nonsense. And, third, if we accept this concept of assumed risk, we lose any meaning that might be given to the concept of *A*'s having made a *commitment* to *B*'s interests over the interests of other person's who might be to some degree worse off than *B* on IC considerations. Surely, our conception of the commitments of friendship, family, and community include a kind of prioritizing which does not include *this* notion of assumed risk, although it might include some other.

Finally, it might be objected that, since it is always permissible for *A* to act according to the IC, and since some R–relationships involve mutual commitments with incurred obligations not to act according to the IC, no such R-relationships are permissible to *A*.

There is nothing paradoxical about this objection, but it does fly in the face of human nature and any plausible view of the kinds of personal projects which make life worth living. There is even some significant question about whether a life without any R-relationships involving the mutual commitments of friendship, family, etc., could

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

be a *personal* life for a human being. Moreover, it most certainly flies in the face of our concept of moral integrity to assert that family and friends and husbands and wives do not have special obligations to each other of the sort in question, because it is *constitutive* of what we take these relationships to be that the parties are committed to each other in these ways.

I conclude, then, that the personal–promise and the personal–relations restrictions are agent-centered restrictions that have as their rationale the preservation of forms of personal integrity allowable under any plausible interpretation of the application of Scheffler's prerogative. Thus, it is paradoxical to accept Scheffler's prerogative departure from consequentialism on any plausible interpretation of its application and to deny these agent–centered restrictions.

GEORGE W. HARRIS

College of William and Mary

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

NEW BOOKS

ADLER, MORTIMER J.: *The Angels and Us.* New York: Macmillan, 1988. ix, 205 p. Paper $6.95.

AIRAKSINEN, TIMO: *Ethics of Coercion and Authority: A Philosophical Study of Social Life.* Pittsburgh: University Press, 1988. ix, 219 p. Cloth $25.95.

ALDRICH, VIRGIL C.: *The Body of a Person.* Lanham, MD: UP of America, 1988. v, 108 p. Cloth $18.95, paper $9.75.

APPELBAUM, RICHARD P.: *Karl Marx.* Newbury Park, CA: Sage, 1988. 160 p. Cloth $19.95, paper $9.95.

BACH, KENT: *Thought and Reference.* New York: Oxford, 1988. ix, 305 p. Cloth $59.00.

BAILEY, JOE: *Pessimism.* New York: Routledge, 1988. ix, 187 p. Cloth $45.00, paper $14.95.

BARNES, ANNETTE: *On Interpretation: A Critical Analysis.* New York: Blackwell, 1988. 171 p. Cloth $45.00, paper $16.95.

BAUMGOLD, DEBORAH: *Hobbes's Political Theory.* New York: Cambridge, 1988. xi, 211 p. Cloth $39.50.

BEETS, M. G. J.: *Reality and Freedom: Reflections on Kant's Moral Philosophy.* Delft: Eburon, 1988. 156 p.

BENN, STANLEY: *A Theory of Freedom.* New York: Cambridge, 1988. xiv, 338 p. Cloth $49.95, paper $17.95.

BENNETT, JONATHAN: *Events and their Names.* Indianapolis: Hackett, 1988. 243 p. Cloth $28.50, paper $16.50.

BERGER, ARTHUR S.: *Evidence of Life After Death: A Casebook for the Tough-minded.* Springfield, IL: Thomas, 1988. x, 147 p. Paper $22.75.

BERMAN, DAVID: *A History of Atheism in Britain: From Hobbes to Russell.* New York: Routledge, 1988. x, 253 p. Cloth $72.50.

BIGELOW, JOHN: *The Reality of Numbers: A Physicalist's Philosophy of Mathematics.* New York: Oxford, 1988. viii, 193 p. Cloth $45.00.

BILLINGTON, RAY: *Living Philosophy: An Introduction to Moral Thought.* New York: Routledge, 1988. xiii, 318 p. Paper $15.50.

BJORK, DANIEL W.: *William James.* New York: Columbia, 1988. xix, 338 p. Cloth $35.00.

BODEN, MARGARET A.: *Problems in the Behavioral Sciences: Computer Models of Mind.* New York: Cambridge, 1988. xi, 289 p. Cloth $42.50, paper $15.95.

BOISVERT, RAYMOND D.: *Dewey's Metaphysics.* New York: Fordham, 1988. xii, 227 p. Cloth $35.00.

BOURNIQUE, GLADYS: *La Philosophie de Josiah Royce.* Paris: Vrin, 1988. ix, 452 p.

BRENNAN, ANDREW: *Conditions of Identity: A Study of Identity and Survival.* New York: Oxford, 1988. ix, 375 p. Cloth $67.00.

BUDZISZEWSKI, J.: *The Nearest Coast of Darkness: A Vindication of the Politics of Virtues.* Ithaca: Cornell, 1988. xi, 201 p. Cloth $24.95.

CAMPBELL, DONALD T.: *Methodology and Epistemology for Social Sciences: Selected Papers.* Chicago: University Press, 1988. xix, 609 p. Cloth $62.50.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

CAVELL, STANLEY: *Themes out of School: Effects and Causes.* Chicago: University Press, 1988. xiv, 268 p. Paper $12.95.

CAWS, PETER: *Structuralism: The Art of the Intelligible.* Atlantic Highlands, NJ: Humanities, 1988. xiii, 276 p. Cloth $45.00.

CHAMPLIN, T. S.: *Reflexive Paradoxes.* New York: Routledge, 1988. viii, 235 p. Cloth $59.95.

CHAPMAN, MICHAEL: *Constructive Evolution: Origins and Development of Piaget's Thought.* New York: Cambridge, 1988. xii, 459 p. Cloth $59.50, paper $22.95.

CHARLTON, WILLIAM: *Weakness of Will: A Philosophical Introduction.* New York: Blackwell, 1988. vii, 196 p. Cloth $45.00, paper $16.95.

CHERRYHOLMES, CLEO H.: *Power and Criticism: Poststructural Investigations in Education.* New York: Teachers College, 1988. xii, 240 p. Cloth $22.95.

CLARK, IAN: *Waging War: A Philosophical Introduction.* New York: Oxford, 1988. 153 p. Cloth $35.00.

COLEMAN, JULES L.: *Markets, Morals, and the Law.* New York: Cambridge, 1988. xvii, 393 p. Cloth $49.00, paper $16.95.

CONNELL, RICHARD J.: *The Empirical Intelligence—The Human Empirical Mode: Philosophy as Originating in Experience.* Lewiston, NY: Mellen, 1988. iii, 285 p. Cloth $49.95.

CONNOLLY, WILLIAM E.: *Political Theory and Modernity.* New York: Blackwell, 1988. xi, 196 p. Cloth $29.95.

COOK, V. J.: *Chomsky's Universal Grammar: An Introduction.* New York: Blackwell, 1988. vi, 201 p. Cloth $45.00, paper $15.95.

COTTINGHAM, JOHN: *The Rationalists.* New York: Oxford, 1988. xii, 234 p. Cloth $29.95, paper $9.95.

CUPITT, DON: *The Sea of Faith: Christianity in Change.* New York: Cambridge, 1988. 286 p. Cloth $29.95.

DAWES, ROBYN M.: *Rational Choice in an Uncertain World.* New York: Jovanovich, 1988. xi, 346 p. Paper $10.00.

DAY, RICHARD B., et al.: *Democratic Theory and Technological Society.* New York: Sharpe, 1988. xii, 356 p. Cloth $39.95.

DEAN, WILLIAM: *History Making History: The New Historicism in American Religious Thought.* Albany: SUNY, 1988. xiv, 175 p. Cloth $34.50, paper $10.95.

DESAN, WILFRED: *Let the Future Come: Perspectives for a Planetary Peace: The Planetary Man: Volume III.* Washington: Georgetown UP, 1988. ix, 164 p. Cloth $17.95, paper $9.95.

DETLEFSEN, MICHAEL: *Hilbert's Program: An Essay on Mathematical Instrumentalism.* Boston: Kluwer, 1988. xiv, 186 p. Cloth $51.00.

DEVRIES, WILLEM A.: *Hegel's Theory of Mental Activity.* Ithaca: Cornell, 1988. xxii, 209 p. Cloth $27.95.

DIEL, PAUL, and JEANNINE SOLOTAREFF: *Symbolism in the Gospel of John.* San Francisco: Harper & Row, 1988. viii, 222 p. Paper $19.95.

DIETZ, MARY G.: *Between the Human and the Divine: The Political Thought of Simone Weil.* Totawa, NJ: Rowman & Littlefield, 1988. xviii, 205 p. Cloth $31.50, paper $14.50.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

DILLON, M. C.: *Merleau-Ponty's Ontology*. Bloomington: Indiana, UP 1988. xi, 286 p. Cloth $35.00.

DILMAN, ILHAM: *Mind, Brain and Behaviour: Discussions of B. F. Skinner and J. R. Searle*. New York: Routledge, 1988. xiv, 142 p. Cloth $49.95.

DRETSKE, FRED: *Explaining Behavior: Reasons in a World of Causes*. Cambridge: MIT, 1988. xi, 165 p. Cloth $22.50.

DUBOIS, PAGE: *Sowing the Body: Psychoanalysis and Ancient Representations of Women*. Chicago: University Press, 1988. xv, 227 p. Cloth $29.95.

EDELSON, MARSHALL: *Psychoanalysis: A Theory in Crisis*. Chicago: University Press, 1988. xxxvi, 392 p. Cloth $39.95.

EKELAND, IVAR: *Mathematics and the Unexpected*. Chicago: University Press, 1988. xiii, 146 p. Cloth $19.95.

EMILSSON, EYJOLFUR K.: *Plotinus on Sense-perception: A Philosophical Study*. New York: Cambridge, 1988. ix, 179 p. Cloth $37.50.

FACKENHEIM, EMIL L.: *What is Judaism?: An Interpretation for the Present Age*. New York: Macmillan, 1988. 320 p. Paper $9.95.

FARRAR, CYNTHIA: *The Origins of Democratic Thinking: The Invention of Politics in Classical Athens*. New York: Cambridge, 1988. x, 301 p. Cloth $37.50.

FEFFER, MELVIN: *Radical Constructionism: Rethinking the Dynamics of Development*. New York: University Press, 1988. 224 p. Cloth $35.00.

FEINBERG, JOEL: *Harmless Wrongdoing: The Moral Limits of the Criminal Law*.New York: Oxford, 1988. xxix, 380 p. Cloth $32.50.

FELEPPA, ROBERT: *Convention, Translation, and Understanding: Philosophical Problems in the Comparative Study of Culture*. Albany: SUNY, 1988. viii, 317 p. Cloth $44.50, paper $14.95.

FINE, ARTHUR: *The Shaky Game: Einstein Realism and the Quantum Theory*. Chicago: University Press, 1988. xi, 186 p. Paper $10.95.

FISHER, ALEC: *The Logic of Real Arguments*. New York: Cambridge, 1988. xi, 190 p. Cloth $34.50, paper $9.95.

FISHER, SIMON: *Revelatory Positivism? Barth's Earliest Theology and the Marburg School*. New York: Oxford, 1988. xv, 348 p. Cloth $69.00.

FITZGIBBONS, ATHOL: *Keynes's Vision: A New Political Economy*. New York: Oxford, 1988. vi, 208 p. Cloth $45.00.

FOGELIN, ROBERT J.: *Figuratively Speaking*. New Haven: Yale, 1988. viii, 120 p. Cloth $10.95.

FOTHERGILL-PAYNE, LOUISE: *Seneca and Celestina*. New York: Cambridge, 1988. xvii, 172 p. Cloth $39.50.

FRANK, ROBERT H.: *Passions with Reason: The Strategic Role of the Emotions*. New York: Norton, 1988. xiii, 304 p. Cloth $19.95.

FRANKFURT, HARRY G.: *The Importance of What We Care About*. New York: Cambridge, 1988. x, 190 p. Cloth $39.50, paper $12.95.

FULLER, STEVE: *Social Epistemology*. Bloomington: Indiana UP, 1988. xv, 316 p. Cloth $27.50.

GÄRDENFORS, PETER: *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. Cambridge: MIT, 1988. xi, 262 p. Cloth $27.50.

GARDINER, PATRICK: *Kierkegaard*. New York: Oxford, 1988. viii, 120 p. Cloth $29.95, paper $5.95.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

GERT, BERNARD: *Morality: A New Justification of the Moral Rules.* New York: Oxford, 1988. xxx, 317 p. Cloth $34.95.

GIBSON, ROGER F. JR.: *Enlightened Empiricism: An Examination of W. V. Quine's Theory of Knowledge.* Gainesville: Florida UP, 1988. xix, 191 p. Cloth $22.50.

GOODMAN, NELSON, and CATHERINE Z. ELGIN: *Reconceptions in Philosophy and Other Arts and Sciences.* Indianapolis: Hackett, 1988. xiv, 174 p. Cloth $21.50.

GORDON, HAIM: *The Other Martin Buber: Recollections of his Contemporaries.* Athens: Ohio UP, 1988. xiii, 186 p. Cloth $18.95.

GRASSI, ERNESTO: *Renaissance Humanism Studies in Philosophy and Poetics.* Albany: SUNY, 1988. xix, 145 p. Cloth $18.00.

GRAYLING, A. C.: *Wittgenstein.* New York: Oxford, 1988. viii, 125 p. Cloth $24.95, paper $4.95.

GREEN, LESLIE: *The Authority of the State.* New York: Oxford, 1988. x, 273 p. Cloth $49.95.

GREENE, MAXINE: *The Dialectic of Freedom.* New York: Teachers College, 1988. xiii, 152 p. Cloth $17.95, paper $9.95.

GRIERE, RONALD N.: *Explaining Science: A Cognitive Approach.* Chicago: University Press, 1988. xix, 321 p. Cloth $34.95.

HAKSAR, VINIT: *Civil Disobedience, Threats and Offers: Gandhi and Rawls.* New York: Oxford, 1988. 58 p. Paper $5.95.

HALE, BOB: *Abstract Objects.* New York: Blackwell, 1988. viii, 282 p. Cloth $55.00.

HALLER, RUDOLF: *Questions on Wittgenstein.* Lincoln: Nebraska UP, 1988. ix, 149 p. Cloth $24.95, paper $9.95.

HANFLING, OSWALD: *The Quest for Meaning.* New York: Blackwell, 1988. xiii, 225 p. Cloth $34.95, paper $16.95.

HARDIN, RUSSELL: *Morality within the Limits of Reason.* Chicago: University Press, 1988. xx, 234 p. Cloth $24.95.

HARRIS, ERROL E.: *The Reality of Time.* Albany: SUNY, 1988. xii, 204 p. Cloth $39.50, paper $12.95.

HART, CLIVE: *Images of Flight.* Berkeley: California UP, 1988. xxvii, 300 p. Cloth $32.50.

HART, W. D.: *The Engines of the Soul.* New York: Cambridge, 1988. xi, 190 p. Cloth $34.50.

HARTAL, PAUL Z.: *The Brush and the Compass: The Interface Dynamics of Art and Science.* Lanham, MD: UP of America, 1988. xxxii, 341 p. Cloth $29.50, paper $17.75.

HARTIGAN, RICHARD S.: *The Future Remembered: An Essay in Biopolitics.* Notre Dame: University Press, 1988. x, 160 p. Cloth $18.95.

HAVELOCK, ERIC A.: *The Muse Learns to Write: Reflections on Orality and Literacy from Antiquity to the Present.* New Haven: Yale, 1988. 144 p. Cloth $20.00, paper $7.95.

HAYES, RICHARD P.: *Dignaga on the Interpretation of Signs.* Boston: Kluwer, 1988. xv, 363 p. Cloth $124.00.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

HEBBLETHWAITE, BRIAN: *The Ocean of Truth: A Defense of Objective Theism*. New York: Cambridge, 1988. x, 165 p. Cloth $34.50, paper $11.95.

HELLER, AGNES: *General Ethics*. New York: Blackwell, 1988. 202 p. Cloth $29.95.

HESS, PETER H.: *Thought and Experience*. Toronto: University Press, 1988. xi, 191 p. Cloth $25.00.

HOLBROOK, DANIEL: *Qualitative Utilitarianism*. Lanham, MD: UP of America, 1988. 141 p. Cloth $22.50, paper $11.25.

HOLIDAY, ANTHONY: *Moral Powers: Normative Necessity in Language and History*. New York: Routledge, 1988. xv, 199 p. Cloth $62.50.

HONDERICH, TED: *A Theory of Determinism: The Mind, Neuroscience, and Life-hopes*. New York: Oxford, 1988. xii, 644 p. Cloth $98.00.

HULL, DAVID L.: *Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science*. Chicago: University Press, 1988. ix, 586 p. $39.95.

HURSTHOUSE, ROSALIND: *Beginning Lives*. New York: Blackwell, 1988. xv, 368 p. Cloth $32.95, paper $16.95.

IDEL, MOISHE: *Kabbalah: New Perspectives*. New Haven: Yale, 1988. xx, 419 p. Cloth $40.00.

ISER, WOLFGANG: *Sterne: Tristam Shandy*. New York: Cambridge, 1988. vii, 135 p. Cloth $19.95, paper $5.95.

JENSON, ROBERT W.: *America's Theologian: A Recommendation of Jonathan Edwards*. New York: Oxford, 1988. xii, 224 p. Cloth $26.00.

JOHNSON, MICHAEL L.: *Mind, Language, Machine: Artificial Intelligence in the Postructuralist Age*. New York: St. Martin's, 1988. xi, 399 p. Cloth $24.95.

KAINZ, HOWARD P.: *Ethics in Context*. Washington: Georgetown UP, 1988. xv, 162 p. Cloth $22.95, paper $11.95.

KALECHOFSKY, ROBERT: *The Persistence of Error: Essays in Developmental Epistemology*. Lanham, MD: UP of America, 1988. v, 73 p. Cloth $17.25, paper $8.75.

KEELEY, MICHAEL: *A Social-contract Theory of Organizations*. Notre Dame: University Press, 1988. x, 288 p. Cloth $32.95.

KEKES, JOHN: *The Unexamined Life*. Cranbury, NJ: Bucknell, 1988. 201 p. Cloth $29.95

KENNEDY, IAN: *Treat Me Right: Essays in Medical Law and Ethics*. New York: Oxford, 1988. xvii, 375 p. Cloth $67.00, paper $24.95.

KERR, FERGUS: *Theology after Wittgenstein*. New York: Blackwell, 1988. xii, 202 p. Cloth $49.95, paper $24.95.

KILCULLEN, JOHN: *Sincerity and Truth: Essays on Arnauld, Bayle, and Toleration*. New York: Oxford, 1988. xii, 228 p. Cloth $49.95.

KITCHING, GAVIN: *Karl Marx and the Philosophy of Praxis*. New York: Routledge, 1988. xi, 265 p. Cloth $49.50, paper $14.95.

KIVY, PETER: *Osmin's Rage: Philosophical Reflections on Opera, Drama, and Text*. Princeton: University Press, 1988. xiii, 303 p. Cloth $29.50.

KUBOVY, MICHAEL: *The Psychology of Perspective and Renaissance Art*. New York: Cambridge, 1988. xiv, 191 p. Paper $16.95.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

LACHS, JOHN: *George Santayana*. Boston: Twayne, 1988. 160 p. Cloth $19.95.

LAMB, DAVID: *Down the Slippery Slope*. New York: Croom Helm, 1988. ix, 134 p. Cloth $45.50.

LEAMAN, OLIVER: *Averoes and his Philosophy*. New York: Oxford, 1988. xii, 204 p. Cloth $49.95.

LEE, SANG HYUNG: *The Philosophical Theology of Jonathan Edwards*. Princeton: University Press, 1988. xiii, 248 p. Cloth $35.00.

LEKAS, PADELIS: *Marx on Classical Antiquity: Problems of Historical Methodology*. New York: St. Martin's, 1988. ix, 286 p. Cloth $35.00.

LEVESQUE-LOPMAN, LOUISE: *Claiming Reality: Phenomenology and Women's Experience*. Totowa, NJ: Rowman & Littlefield, 1988. xix, 188 p. Cloth $29.50, paper $13.95.

LEVIN, SAMUEL R.: *Metaphoric Worlds: Conceptions of Romantic Nature*. New Haven: Yale, 1988. xv, 251 p. Cloth $26.50.

LEVINSON, PAUL: *Mind at Large: Knowing in the Technological Age*. Greenwich, CT: Jai, 1988. xviii, 271 p. Cloth $31.75.

LIEBERSOHN, HARRY: *Fate and Utopia in German Sociology, 1870–1923*. Cambridge: MIT, 1988. x, 282 p. Cloth $27.50.

LOEWALD, HANS W.: *Sublimation: Inquiries into Theoretical Psychoanalysis*. New Haven: Yale, 1988. ix, 88 p. Cloth $15.95.

LOY, DAVID: *Nonduality: A Study in Comparative Philosophy*. New Haven: Yale, 1988. ix, 346 p. Cloth $32.50.

LUNTLEY, MICHAEL: *Language, Logic and Experience: The Case for Anti-Realism*. Peru, IL: Open Court, 1988. x, 278 p. Cloth $29.95, paper $14.95.

LUTZ, CATHERINE A.: *Unnatural Emotions: Everyday Sentiments on a Micronesian Atoll and their Challenge to Western Theory*. Chicago: University Press, 1988. xii, 273 p. Cloth $35.00, paper $13.95.

MARX, GARY T.: *Under Cover: Police Surveillance in America*. Berkeley: California UP, 1988. xxv, 283 p. Cloth $25.00.

MASON, J. K.: *Human Life and Medical Practice*. Edinburgh: University Press, 1988. viii, 161 p.

MCCARTHY, GEORGE E.: *Marx's Critique of Science and Positivism: The Methodological Foundations of Political Economy*. Boston: Kluwer, 1988. xi, 225 p. Cloth $74.00.

MCCORMICK, PETER J.: *Fictions, Philosophies, and the Problems of Poetics*. Ithaca: Cornell, 1988. xvi, 351 p. Cloth $42.50, paper $15.95.

MCGUINNESS, BRIAN: *Wittgenstein: A Life: Young Ludwig: 1889–1921*. Berkeley: California, 1988. xii, 322 p. Cloth $22.50.

MCLUHAN, MARSHALL, and ERIC MCLUHAN: *Laws of Media: The New Science*. Toronto: University Press, 1988. xi, 252 p. Cloth $27.50.

MCPHAIL, ROBYN, and DAVID E. WARD: *Morality and Agency*. Lanham, MD: UP of America, 1988. xviii, 165 p. Cloth $28.50, paper $12.75.

MEEHAN, EUGENE J.: *The Thinking Game*. Chatham, NJ: Chatham, 1988. ix, 244 p. Paper $12.95.

MELDEN, A. I.: *Rights in Moral Lives: A Historical-Philosophical Essay*. Berkeley: California UP, 1988. x, 155 p. Cloth $25.00.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

MELLING, DAVID J.: *Understanding Plato.* New York: Oxford, 1987. vi, 178 p. Cloth $32.00, paper $9.95.

MEYER, ROBERT G.: *The Likelihood of Knowledge.* Boston: Kluwer, 1988. xi, 188 p. Cloth $49.00.

MITIAS, MICHAEL H.: *What Makes an Experience Aesthetic?* Amsterdam: Rodopi, 1988. 154 p.

MOHR, RICHARD D.: *Gays/Justice: A Study of Ethics, Society, and Law.* New York: Columbia, 1988. ix, 357 p. Cloth $30.00.

MOKED, GABRIEL: *Particles and Ideas: Bishop Berkeley's Corpuscularian Philosophy.* New York: Oxford, 1988. viii, 245 p. Cloth $55.00.

MYERS, GERALD E.: *William James: His Life and Thought.* New Haven: Yale, 1988. xv, 628 p. Paper $19.95.

NEWMEYER, FREDERICK J.: *The Politics of Linguistics.* Chicago: University Press, 1988. vii, 171 p. Paper $9.95.

NEWSOM, ROBERT: *A Likely Story: Probability and Play in Fiction.* New Brunswick, NJ: Rutgers, 1988. 213 p. Cloth $30.00, paper $15.00.

NYE, ANDREA: *Feminist Theory and the Philosophies of Man.* New York: Routledge, 1988. x, 244 p. Cloth $49.95.

O'CONNOR, BERNARD F.: *A Dialogue between Philosophy and Religion: The Perspective of Karl Jaspers.* Lanham, MD: UP of America, 1988. xviii, 209 p. Cloth $26.25, paper $12.75.

O'HEAR, ANTHONY: *The Element of Fire: Science, Art, and the Human World.* New York: Routledge, 1988. xi, 178 p. Cloth $34.50.

OLEN, JEFFREY: *Moral Freedom.* Philadelphia: Temple, 1988. xi, 151 p. Cloth $19.95.

ORTONY, ANDREW, GERALD L. CLORE, and ALLAN COLLINS: *The Cognitive Structure of Emotions.* New York: Cambridge, 1988. xi, 207 p. Cloth $29.95.

PATOCKA, JAN: *Le Monde Naturel et le mouvement de l'Existence Humaine.* Boston: Kluwer, 1988. xxiii, 276 p. Cloth $89.00.

PAULSON, MICHAEL G.: *The Possible Influence of Montaigne's* Essais *on Descartes'* Treatise on the Passions. Lanham, MD: UP of America, 1988. xiii, 123 p. Cloth $22.50, paper $10.25.

PEARL, JUDEA: *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Mountain View, CA: Kaufmann, 1988. xix, 552 p. Cloth $39.95.

PERRY, MICHAEL J.: *Morality, Politics, and Law.* New York: Oxford, 1988. 323 p. Cloth $29.95.

PFEFFERKORN, KRISTIN: *Novalis: A Romantic's Theory of Language and Poetry.* New Haven: Yale, 1988. xii, 308 p. Cloth $30.00.

PHILLIPS, D. Z.: *Faith after Foundationalism.* New York: Routledge, 1988. xviii, 341 p. Cloth $49.95.

PINKARD, TERRY: *Hegel's Dialectic: The Explanation of Possibility.* Philadelphia: Temple, 1988. xi, 236 p. Cloth $34.95.

PUTNAM, HILARY: *Representation and Reality.* Cambridge: MIT Press, 1988. xv, 136 p. Cloth $19.95.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

QUINN, DENNIS P.: *An Examination of Kant's Treatment of Transcendental Freedom.* Lanham, MD: UP of America, 1988. x, 90 p. Cloth $16.50, paper $8.25.

REED, EDWARD S.: *James J. Gibson and the Psychology of Perception.* New Haven: Yale, 1989. ix, 348 p. Cloth $32.50.

REEVE, C. D. C.: *Philosopher-Kings: The Argument of Plato's* Republic. Princeton: University Press, 1989. xv, 350 p. Cloth $35.00.

REEVES, M. FRANCIS: *Problems of Philosophy and Society: A Conversation with Plato.* Lanham, MD: UP of America, 1988. xii, 134 p. Paper $9.25.

REISS, TIMOTHY J.: *The Uncertainty of Analysis: Problems in Truth, Meaning, and Culture.* Ithaca: Cornell, 1988. xiii, 298 p. Cloth $29.95.

RICH, DAVID Z.: *The Dynamics of Knowledge: A Contemporary View.* New York: Greenwood, 1988. viii, 240 p. Cloth $39.95.

RICKMAN, H. P.: *Dilthey Today.* New York: Greenwood, 1988. xvii, 194 p. Cloth $39.95.

RILEY, JONATHON: *Liberal Utilitarianism.* New York: Cambridge, 1988. 398 p. Cloth $39.50.

ROBINSON, WILLIAM S.: *Brains and People: An Essay on Mentality and its Causal Conditions.* Philadelphia: Temple, 1988. xiv, 225 p. Cloth $34.95.

RORTY, AMELIE OKSENBERG: *Mind in Action: Essays in the Philosophy of Mind.* Boston: Beacon, 1988. x, 378 p. Cloth $27.50.

ROTH, MICHAEL S.: *Knowing and History: Appropriations of Hegel in Twentieth-century France.* Ithaca: Cornell, 1988. xv, 264 p. Cloth $28.50.

ROWE, WILLIAM L., and WILLIAM J. WAINWRIGHT.: *Philosophy of Religion.* San Diego: Harcourt, 1988. xv, 491 p. Paper $14.00.

RUSE, MICHAEL: *Philosophy of Biology Today.* Albany: SUNY, 1988. x, 155 p. Cloth $29.00, paper $9.95.

SAINSBURY, R. M.: *Paradoxes.* New York: Cambridge, 1988. vii, 163 p. Paper $7.95.

SAYEGH, NAWAL SARRAF: *Philosophy and Philosophers: From Thales to Tillich.* Ottawa: Academy, 1988. xi, 290 p.

SCHLESINGER, GEORGE N.: *New Perspectives on Old Time Religion.* New York: Oxford, 1988. 196 p. Cloth $34.50.

SCHOTT, ROBIN MAY: *Cognition and Eros: A Critique of the Kantian Paradigm.* Boston: Beacon, 1988. xiv, 247 p. Cloth $21.95.

SCHRAG, FRANCIS: *Thinking in School and Society.* New York: Routledge, 1988. vii, 189 p. Cloth $25.00.

SCOLNICOV, SAMUEL: *Plato's Metaphysics of Education.* New York: Routledge, 1988. viii, 157 p. Cloth 27.50.

SEESKIN, KENNETH: *Dialogue and Discovery: A Study in Socratic Method.* Albany: SUNY, 1987. x, 179 p. Cloth $44.50, paper $14.95.

SEPPER, DENNIS L.: *Goethe contra Newton.* New York: Cambridge, 1988. xvi, 222 p. Cloth $39.50.

SICKER, MARTIN: *The Judaic State: A Study in Rabbinic Political Theory.* New York: Praeger, 1988. 161 p. Cloth $37.95.

SIEGAL, HARVEY: *Educating Reason: Rationality, Critical Thinking, and Education.* New York: Routledge, 1988. xiii, 191 p. Cloth $25.00.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

SIMMONS, JOE: *Vision and Spirit: An Essay on Plato's Warrior Class.* Lanham, MD: UP of America, 1988. v, 55 p. Cloth $16.50, paper $8.25.

SINGH, CHHATRAPATI: *Law from Anarchy to Utopia.* New York: Oxford, 1988. xviii, 299 p. Cloth $65.00.

SINNOTT-ARMSTRONG, WALTER: *Moral Dilemmas.* New York: Blackwell, 1988. viii, 264 p. Cloth $34.95.

SMITH, BARBARA HERRNSTEIN: *Contingencies of Value: Alternative Perspectives for Critical Theory.* Cambridge: Harvard, 1988. 229 p. Cloth $22.50.

SMITH, JOHN H.: *The Spirit and its Letter: Traces of Rhetoric in Hegel's Philosophy.* Ithaca: Cornell, 1988. xii, 283 p. Cloth $32.95.

SMITH, STEPHEN G.: *The Concept of the Spiritual: An Essay in First Philosophy.* Philadelphia: Temple, 1988. xiv, 308 p. Cloth $34.95.

SORABJI, RICHARD: *Matter, Space, and Motion: Theories in Antiquity and Their Sequel.* Ithaca: Cornell, 1988. x, 377 p. Cloth $49.50.

SORENSEN, ROY A.: *Blindspots.* New York: Oxford, 1988. xiii, 456 p. Cloth $72.00.

SPARSHOT, FRANCIS: *Off the Ground: First Steps to a Philosophical Consideration of the Dance.* Princeton: University Press, 1988. xxiii, 430 p. Cloth $49.50.

STAAL, FRITZ: *Universals: Studies in Indian Logic and Linguistics.* Chicago: University Press, 1988. 267 p. Cloth $47.50, paper $18.95.

STENGER, VICTOR J.: *Not by Design: The Origin of the Universe.* Buffalo: Prometheus, 1988. 202 p. Cloth $22.95.

STERBA, JAMES P.: *How to Make People Just: A Practical Reconciliation of Alternative Conceptions of Justice.* Totowa, NJ: Rowman & Littlefield, 1988. vii, 200 p. Cloth $33.00, paper $14.95.

STOJANOVIC, SVETOZAR: *Perestroika: From Marxism and Bolshevism to Gorbachev.* Buffalo: Prometheus, 1988. 167 p. Cloth $23.95.

STOUT, JEFFREY: *Ethics after Babel: The Languages of Morals and their Discontents.* Boston: Beacon, 1988. xiv, 338 p. Cloth $27.50.

SUKALE, MICHAEL: *Denken, Sprechen und Wissen.* Tübingen: Mohr, 1988. xiii, 325 p.

SWEET, ALBERT M.: *The Pragmatics and Semiotics of Standard Languages.* University Park: Pennsylvania State UP, 1988. 209 p. Cloth $24.95.

TESON, FERNANDO R.: *Humanitarian Intervention: An Inquiry into Law and Morality.* Ardsley-on-Hudson, NY: Transnational, 1988. xv, 272 p. Cloth $50.00.

THAGARD, PAUL: *Computational Philosophy of Science.* Cambridge: MIT, 1988. xiv, 240 p. Cloth $25.00.

THOMAS, D. A. LLOYD: *In Defense of Liberalism.* New York: Blackwell, 1988. ix, 133 p. Cloth $39.95.

VAN DER SANDT, ROB A.: *Context and Presuppostion.* New York: Croom Helm, 1988. xvii, 249 p. Cloth $65.00.

VAN BENTHEM, JOHAN: *Modal Logic and Classical Logic.* Atlantic Highlands, NJ: Humanities, 1985. 234 p. Cloth $50.00.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

VJECSNER, PAUL: *On Proof for the Existence of God, and Other Reflective Inquiries.* New York: Penden, 1988. 258 p. Cloth $20.00.

WALLACE, JAMES D.: *Moral Relevance and Moral Conflict.* Ithaca: Cornell, 1988. x, 167 p. Cloth $19.95.

WARREN, MARK: *Nietzsche and Political Thought.* Cambridge: MIT, 1988. xv, 311 p. Cloth $25.00.

WEDIN, MICHAEL V.: *Mind and Imagination in Aristotle.* New Haven: Yale, 1988. xiv, 292 p. Cloth $30.00.

WEITZ, MORRIS: *Theories of Concepts: A History of the Major Philosophical Tradition.* New York: Routledge, 1988. xxi, 310 p. Cloth $38.00.

WHITE, ALAN R.: *Methods of Metaphysics.* New York: Routledge, 1988. 203 p. Cloth $49.95.

WHITESIDE, KERRY H.: *Merleau-Ponty and the Foundation of an Existential Politics.* Princeton: University Press, 1988. ix, 339 p. Cloth $40.00, paper $16.00.

WHITE, STEPHEN K.: *The Recent Work of Jürgen Habermas: Reason, Justice and Modernity.* New York: Cambridge, 1988. xi, 190 p. Cloth $29.95.

WILKES, KATHLEEN V.: *Real People: Personal Identity without Thought Experiments.* New York: Oxford, 1988. x, 249 p. Cloth $49.95.

WILL, FREDERICK L.: *Beyond Deduction: Ampliative Aspects of Philosophical Reflection.* New York: Routledge, 1988. x, 260 p. Cloth $32.50.

WINFIELD, RICHARD DIEN: *Reason and Justice.* Albany: SUNY, 1988. xii, 318 p. Cloth $49.50, paper $18.95.

WINTERBOURNE, A. T.: *The Ideal and the Real: An Outline of Kant's Theory of Space, Time and Mathematical Construction.* Boston: Kluwer, 1988. xi, 137 p. Cloth $50.00.

WOODWARD, ANTHONY: *Living in the Eternal: A Study of George Santayana.* Nashville: Vanderbilt UP, 1988. ix, 171 p. Cloth $22.50.

ZALTA, EDWARD N.: *Intensional Logic and the Metaphysics of Intentionality.* Cambridge: MIT, 1988. xiii, 256 p. Cloth $25.00.

ZUCKERT, CATHERINE H.: *Understanding the Political Spirit.* New Haven: Yale, 1988. viii, 203 p. Cloth $21.50.

(1) The following are the correct deadlines for the submission of Johnsonian Prize manuscripts:

AUGUST 1, 1989 FOR THE 1989 PRIZE
JANUARY 15, 1990 FOR THE 1990 PRIZE

We apologize for the incorrect dates published in our January issue.
(2) In our December issue, page 723, the name of Professor James Gutmann was regrettably spelled wrong. We would like to apologize to all those who knew him, and remind you of The James Gutmann Memorial Fund at Columbia University.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVI, NO. 3, MARCH 1989

QUINE'S POINT OF VIEW*

There is a richness and complexity, as well as much apparent paradox, in W. V. O. Quine's writings which make understanding his arguments difficult and misunderstandings of them frequent. Simple misunderstandings have led to unconvincing criticisms of his views. Richard Schuldenfrei[1] was one of the first to urge that a proper understanding of the texts is a necessary preliminary to evaluating the arguments. In particular, Shuldenfrei argued, it is important to spell out the point of view guiding the arguments. We need to know what Quine means when he claims that methods of translation are not, and cannot be made, objective: we need to know what is meant by 'objective' in such claims. Only then can there be a fair appraisal of his views.

No consensus has been reached on the nature of Quine's point of view. Schuldenfrei claims that Quine is doing science and writing from within a scientific theory of the world. Hilary Putnam[2] argues that Quine is working from within our ordinary conceptual scheme. Peter Hylton,[3] Joseph Levine,[4] Thomas Ricketts,[5] and the recent

* I am grateful to Burton Dreben, Warren Goldfarb, Donald Gustafson, James Higginbotham, Mark Kaplan, Joseph Levine, Ted Morris, Nickolas Pappas, Hilary Putnam, Georges Rey, Robert Richardson, and Marya Schechtman for helpful discussions on the issues covered in this paper. This work was partly supported by a Summer Faculty Research Fellowship (1987) from the University Research Council, University of Cincinnati.

[1] "Quine in Perspective," this JOURNAL, LXIX, 1 (January 13, 1972): 5–16.

[2] "The Refutation of Conventionalism," in *Mind, Language and Reality* (New York: Cambridge, 1975). Putnam argues here that Quine$_2$, who investigates how well defined the word 'translation' is, is the real Quine.

[3] "Analyticity and the Indeterminacy of Translation," *Synthese*, LII (1982): 167–184.

[4] "Quine on Psychology," in *Naturalistic Epistemology*, Abner Shimony and Debra Nails, eds. (Boston: Reidel, 1987), pp. 259–290.

[5] "Rationality, Translation and Epistemology Naturalized," this JOURNAL, LXXIX, 3 (March 1982): 117–136.

0022-362X/89/8603/113–136 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Putnam[6] argue that Quine's conclusions rest on his taking physics as the ultimate parameter of reality. And Noam Chomsky[7] and Michael Friedman[8] hold that the arguments for indeterminacy rest on behaviorist presuppositions. Some of these differences in interpretation have to do with differences in the critics' own philosophical views, and some with their emphasizing what they take to be the best, or most promising, arguments in Quine's work. But some of the differences also stem from problems with understanding the texts. Quine himself characterizes the point of view from which he argues for indeterminacy differently on different occasions: sometimes he says that indeterminacy is a consequence of his behaviorism,[9] sometimes a consequence of his physicalism,[10] sometimes merely a consequence of reflection on the available data and methods for translation,[11] and sometimes a consequence of the underdetermination of physical (not semantic) theory.[12] Along with this, the arguments for indeterminacy are different—even contrary—on different occasions. Is Quine's work motivated by more than one point of view? If so, what does each involve, and how are they related to one another? In particular, how can Quine's repeated and attractive claim to be "working from within"[13] our ordinary conceptual scheme in the direction of scientific advance be reconciled with the startlingly counterintuitive doctrine of indeterminacy and the less innocent commitments of doctrines such as physicalism and behaviorism? Prior interpretations of Quine's work have not answered these questions.

[6] "A Comparison of Something with Something Else," *New Literary History*, XVII (1985/6): 61–69.

[7] "Quine's Empirical Assumptions," in *Words and Objections*, Donald Davidson and Jaakko Hintikka, eds. (Boston: Reidel, 1969), pp. 53–68.

[8] "Physicalism and the Indeterminacy of Translation," *Nous*, IX (1975): 353–374.

[9] For example, in *Word and Object* (Cambridge: MIT, 1960) [henceforth WO]; "Indeterminacy of Translation Again," this JOURNAL, LXXXIV, 1 (January 1987), p. 5.

[10] This is not a view espoused in WO (apart from the rather unclear passage on pp. 76/7). An example of the view is in the "Reply to Chomsky," *Words and Objections*, p. 303.

[11] This is the view taken in the bulk of WO, ch. 2, where available data and methods are examined and found insufficient to determine translation. It is often repeated, e.g., in the "Reply to Hintikka," in *Words and Objections*, p. 314; and, e.g., in "Facts of the Matter," in *American Philosophy from Edwards to Quine*, R. W. Shahan and K. R. Merrill, eds. (Norman: Oklahoma UP, 1977), where he writes on p. 193 that meanings are indeterminate because we have no clear way to handle them.

[12] For example, he suggests this at the beginning and end of WO, ch. 2 (pp. 26, 78) and at the beginning of "On the Reasons for Indeterminacy of Translation," this JOURNAL, LXVII, 3 (March 1970), p. 179. I shall explain this suggestion in section II of this paper.

[13] The phrase comes from "Posits and Reality" (1955), in *Ways of Paradox* (Cambridge: Harvard, 1976), p. 252. The claim to be "working from within" is

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The aim of this paper is to close the debate over what point of view underlies Quine's work. I do so by providing a new interpretation of the texts. I argue that there is a single point of view, which I call *natural empiricism,* underlying and generating all of Quine's doctrines: from the specific claims regarding indeterminacy, underdetermination, and truth, to the more general claims to be physicalist, behaviorist, empiricist, and "working from within." One of the virtues of this interpretation is that it shows the systematic nature of Quine's thought; another is that it will become clear why, and how, indeterminacy is so central a doctrine for Quine.

My strategy in arguing for this interpretation is as follows. In sections I and II, I set out different arguments and positions that Quine puts forward, mostly on the topic of indeterminacy. In section III, based on the results of sections I and II, I present my interpretation of Quine's views and show its explanatory force. Finally, in section IV, I suggest how this interpretation should be used in understanding and evaluating Quine's views.

I. FIRST ARGUMENT FOR INDETERMINACY

I discern two different strategies in the arguments for indeterminacy. The first is more prominent in the texts than the second. Perhaps because of this, the secondary literature focuses on it.[14] It involves reflection on the methodology of translation, often through imagining the situation of radical translation, where the field linguist is trying to translate the native language into his or her own language without the aid of interpreters, dictionaries, and so forth. The idea is that we find out what is involved in translation through imagining a situation in which nothing about the meanings of the native utterances is taken for granted. WO, chapter 2, which is Quine's first full statement of indeterminacy, is mostly concerned with pursuing this strategy. I begin by explaining this. Then I shall show how Quine subsequently changes and vacillates on what he thinks are the results of the argument.

In WO, chapter 2, Quine raises skeptical questions regarding our translational practices and maintains that, using all the data and methods we can muster, we cannot perform determinate translation for all the sentences of a language. The extent of determinate translation, according to this argument, is the (holophrastically construed) observation sentences and truth functional connectives (cf.

---

central in Quine's work; so central that Neurath's metaphor of the need to repair one's boat while afloat forms the epigraph for the whole of WO.

[14] An exception I have found is in Levine, "Quine on Psychology," which sees Quine as arguing in two ways, and bases this interpretation on some of the same texts that I do. We differ on precisely what these arguments are, and on our assessments of them.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

WO, 68). Also, the objective part of the practice of translation involves preservation of the truth value of as many obvious statements as possible—i.e., the principle of charity.[15] And, possibly, simpler translations are preferred.[16] The rest of translation, Quine claims, proceeds via subjective choices. The gavagai example takes center stage as the suggestion that it is inscrutable whether the native talks of rabbits, rabbit parts, rabbit time slices, rabbit stages, or something else in the rabbit situation.

In this argument from the methodology of translation, Quine is arguing that the available data for translation are behavioral facts, and that these data (along with additional constraints such as charity and simplicity) fail to determine unique choice of translation manual. The general character of the argument is to consider the actual situation of radical translation and assess how much of the method used is objective. The argument has a number of difficulties, some of which Quine has commented on in the years since WO.

One difficulty concerns the skeptical style of the argument. This argument comes under pressure in two directions: to yield more or less skeptical results. If Quine's skeptical considerations are applied consistently, they yield the result that *none* of language is determinately translatable, not even the signs for assent and dissent which are required for translation of observation sentences. For one could ask the skeptical question, how can we tell that the native 'evet' should be translated by our 'yes', rather than by our 'no' or by some other English utterance? Unless we start by making assumptions about native linguistic behavior, such as that the native will assent when mimicked and that the native is generally veracious, and so forth, translation cannot even begin. But what could be the evidence for such assumptions? Once, in reply to some similar suggestions by Jaakko Hintikka, Quine allowed that all translation is indeterminate.[17] But, since that reply, he has gone back to claiming that trans-

[15] See WO, 68. In WO, Quine seems to adopt the principle of charity as an objective constraint on translation to be applied quite generally. That is, it seems most reasonable not to regard the native as holding blatantly false beliefs. The principle of charity is involved in translation not only for translation of the truth functional connectives (WO, 59) but also more generally for translation of stimulus analytic sentences (WO, 68). Harman pointed out to Quine (private correspondence, April 6, 1970) that, if translation of truth functions is determinate, Quine is committed to regarding the principle of charity as objective generally, that is, for translation of stimulus analytic sentences also.

[16] See WO, 69, where Quine talks of preference for "neat" analytical hypotheses. I read this as an objective preference, since Quine's suggestion is that it be balanced with the principle of charity, and he seems to regard the principle of charity as objective. It is only on p. 70 that he begins to consider subjective maxims guiding translation.

[17] "Reply to Hintikka," in *Words and Objections*, p. 312.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

lation of assent and dissent is determinate,[18] although his arguments for this are susceptible of the same skeptical attack. It is not surprising that Quine wants to maintain that translation of assent and dissent is determinate, for without their determinate translation, one cannot even translate observation sentences, and without determinate translation of observation sentences, there is no fact of the matter about empirical content of theories. As I shall show later, Quine's insistence that there is a fact of the matter about empirical content of theories[19] is an indication of his underlying point of view. So, if Quine's skeptical considerations are pressed consistently, they yield results that he is unwilling to accept.

There is also pressure to move in the other direction and dispel the skeptical suggestions. I claim that his construal of the available behavioral data is too narrow, and his suggestions regarding the subjectivity of our translation practices unconvincing. These claims are developed in the next three paragraphs.

The data for translation which he considers are words said in response to other words, often in the presence of sensory stimulation. This leaves out judgments based on behavior other than phoneme production. For example, one speaker may encourage and reinforce some particular speech pattern of another, and this behavior should be taken as constitutive of assent. Tracking behavior and looks of recognition should be taken as indications of what the speaker is prompted to talk about. And counting behavior can aid translation of terms. Quine regards judgments based on such behavioral data as "intuitive" (WO, 30). By this he means an account in which "terms are used in habitual ways, without reflecting on how they might be defined or what presuppositions they may conceal" (WO, 36, fn.): in other words, "intuitive" judgments do not result in determinate translation, since they may involve concealed analytical hypotheses. I think the difficulty here is not the circularity of "intuitive" judgments, however, but the complexity that they involve. The behavior typical of assent seems no less complex than the behavior typical of counting. Spelling out the behavioral data that we use to perform translation in a non-question-begging way is a lengthy project which I cannot pursue here,[20] but there is no reason to think it will fail.

[18] For example, he argues that assent and dissent are scrutable in "Mind and Verbal Dispositions," in *Mind and Language: Wolfson College Lectures 1974*, Samuel Guttenplan, ed. (New York: Oxford, 1975), p. 91.

[19] See, for example, Quine's recent remarks in *The Philosophy of W. V. Quine*, L. E. Hahn and P. A. Schillp, eds. (La Salle, IL: Open Court, 1986), pp. 402, 427.

[20] I say more in an unpublished doctoral dissertation, *Quine's Point of View* (Harvard, 1986). Note that when I write in terms of using behavioral data to

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

There are other reasons to suspect that Quine is being overly skeptical of our translation practices. He claims that we settle on a single translation manual by the use of subjective maxims and even random choices,[21] and that both ordinarily go unnoticed by us. The claim is that, in using subjective maxims, we tend to project our own conceptual scheme into the native's language (e.g., WO, 70). Quine does not, however, give convincing examples of projection or random choice. He gives no examples of randomness, and, indeed, the similar results of separate enterprises of translation argue against this. And the examples he gives of projection seem to show that we are aware of conceptual differences and either do not attempt translation or regard the results as approximate. To respond to the example of the Japanese classifiers in "Ontological Relativity"[22]: we recognize that there is no perfect translation into English and translate approximately in one of two ways. Perhaps all translation is more or less approximate, but this does not make the maxims used subjective. The example of translation of our physical theory into the native language (WO, 76) is said to make indeterminacy "visible." What is made visible is that there are concepts expressible in English which are not expressible in the native language. But this is not visible indeterminacy. Rather, the case exhibits our unwillingness to translate by projecting our conceptual scheme into the native's. It shows our awareness of conceptual differences between languages. It is a case of visible conceptual difference rather than visible inscrutability or visible projection.

Finally, Quine is overskeptical when he takes the view (as he sometimes does) that there is determinate translation only if there are "linguistically neutral meanings,"[23] that is, only if there are interculturally shared meanings. Although this is a simple and traditional

---

translate, I do not mean that we must do so in terms of the trajectories of the native's moving parts (e.g., limbs, head, vocal cords). This is for two reasons. First, words describing behavior are more coarse-grained. Just as we do not have to describe the trajectory of each particle of a sailcloth when we describe the behavior of the mainsail as "luffing," we do not need to give a micro-level description of what is happening with the native. Secondly, as Fred Dretske has pointed out [*Explaining Behavior* (Cambridge: MIT, 1988), ch. 1], our notion of behavior involves the causal origin of the muscle movements. Neither of these reasons suggests that ordinary descriptions of behavior are circular in the sense that they presume analytical hypotheses.

[21] See WO, 72, where Quine writes: "Using what first comes to mind engenders an air of determinacy though freedom reign."

[22] In *Ontological Relativity and Other Essays* (New York: Columbia, 1969), p. 35.

[23] For example, WO, 76, and the passage cited above in "Facts of the Matter" (fn. 11).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

view of the conditions for determinate translation, one in which "linguistically neutral meanings" ground our use of words, such an account—the propositional model—has not been shown to be necessary for the vindication of the objectivity of our practices of ordinary translation or the meaningfulness of ordinary discourse. Indeterminacy is a stronger thesis than the rejection of the propositional model. Translations can be more or less approximate, depending on one's purposes, so that one talks loosely of "sameness of meaning" without there being the identities of meaning required on the propositional model. The approximations made in translation are not a basis for regarding our methods of translation as subjective.

Thus, the skeptical considerations Quine uses to argue for indeterminacy are unconvincing. If the lack of a full account of objective methods for translation is sufficient to justify skeptical conclusions, then the skepticism can be pushed to an extreme that even Quine is unwilling to accept. If not, then it is not clear that the argument can get off the ground at all. Quine vacillates on how much indeterminacy the argument reveals, once denying that translation of assent and dissent is determinate (cf. fn. 17), other times going along with maximum application of the principle of charity and any other available maxims.[24] It is hard to believe that these skeptical considerations are what is primarily motivating Quine in arguing for indeterminacy; it would not be surprising to find out that his conviction that there is indeterminacy arises from another source.

Before turning to the second argument for indeterminacy, I shall discuss a further difficulty with the first argument. This is a difficulty that Chomsky and others have raised. In WO, chapter 1, and elsewhere, Quine claims that there is underdetermination of physical theory. The underdetermination is not merely a matter of there being more than one theory generating correct predictions. For Quine claims that, even if other canons of scientific method such as simplicity and conservatism are included, there is still underdetermination.[25] Yet Quine maintains that this underdetermination should not make us skeptical of the truth of our own theory. Inspired by Alfred Tarski's work, he adopts the view that there is no more to claiming that a theory is true than to adopting the theory. This account of truth allows Quine to hold that our ordinary and scientific views of the world are true, despite the possibility of equally successful competitors. It also renders Quine's first argument for

[24] For example, see "Comments on Newton-Smith," *Analysis*, XXXIX (March 1979), pp. 66/7.

[25] See WO, 23, where Quine writes: "Scientific method is the way to truth, but it affords even in principle no unique definition of truth."

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

indeterminacy incomplete. For, if there are, as Quine argues, a number of translations in agreement with the behavioral data and satisfying all other objective constraints, then why not say that the translation that *we* adopt gives the true meanings?

In WO, Quine raises this question, but the answer he gives there is unsatisfactory.[26] Later, prompted by his critics, he writes that the assumption of physicalism completes the argument for indeterminacy.[27] Physicalism is the view that physics gives us a complete theory of the world; so that, once a physical theory is chosen, all the facts are settled. Fixing the physical facts, however, will not decide between analytical hypotheses that satisfy all objective constraints on translation. So, speaking from the point of view of physics, translation is indeterminate.

It has been concluded from this that physicalism is at the heart of Quine's arguments for indeterminacy.[28] Quine denies this, however:

> . . . I want to relate physicalism to my perennial criticisms of mentalistic semantics. Readers have supposed that my complaint is ontological; it is not. If in general I could make satisfactory sense of declaring two expressions to be synonymous, I would be more than pleased to recognize an abstract object as their common meaning. The method is familiar: I would define the meaning of an expression as the set of its synonyms. Where the trouble lies, rather, is in the two-place predicate of synonymy itself; it is too desperately wanting in clarity and perspicuity ("Facts of the Matter," p. 193).

So, although Quine agrees that he is a physicalist, he denies that this is at the root of his complaints about discussions of meaning. Instead, he claims that talk of meanings is simply unclear. This brings the discussion back to the considerations used in the first argument for indeterminacy.[29]

A way to understand this apparent inconsistency in Quine's writing is as follows. If the first argument for indeterminacy shows what Quine claims—that a number of translation manuals are in agreement with behavioral data and objective constraints—then, taken out of the context of Quine's views on truth, it would be a complete

[26] See WO, 76, where Quine claims that we have to accept a theory of the physical world, but we do not have to accept a translation manual. This is unsatisfactory, because, surely, in using the translation manual, we are accepting it, and we need a translation manual in order to translate.

[27] See "Reply to Chomsky."

[28] See, for example, the views of Hylton, Levine, Ricketts, and Putnam cited above.

[29] In this paper, as elsewhere, Quine treats the question of indeterminacy and the question of getting a mentalistic semantics interchangeably. I commented on this above.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

argument. We are, ordinarily, sensitive to underdetermination claims, whether in physics or translation. If more than one theory is successful, then ordinarily we would say that neither can lay claim to truth. Quine's argument becomes incomplete only because he has out-of-the-ordinary views about truth. I shall claim that his views on truth come from an aspect of his thinking which is not present in the first argument for indeterminacy and is in tension with it. This is an example of the tension.

So, the problem is one of internal consistency of Quine's views. If the first argument is taken on its own and is successful in showing the existence of alternative translation manuals, then it is a serious argument for us. I think it is because Quine recognizes this, and also because the considerations he employs in discussing truth are not uppermost in the first argument for indeterminacy, that he denies that physicalism is at the root of the argument. When pushed on the internal inconsistency of his views, however, Quine grants that physicalism is required.

To summarize the discussion of indeterminacy so far: when arguing that the data and methods for translation do not suffice to determine choice of analytical hypotheses, Quine uses skeptical considerations that are not persuasive. The argument is unstable; Quine has vacillated on how seriously to take the skeptical considerations. And I gave some indication of how one might counter the skeptical suggestions. It seems that this argument alone cannot be the source of Quine's conviction that there is indeterminacy. Furthermore, the argument, even if it succeeded on its own terms, is incomplete if taken in the context of Quine's views on underdetermination and truth. The assumption of physicalism completes the argument. I argued, with Quine, that this does not mean that physicalism is at the root of his first argument for indeterminacy. It remains to be seen where it belongs in other parts of his work and what relation it has to his unusual views on underdetermination and truth.

II. SECOND ARGUMENT FOR INDETERMINACY

Quine has a second argument for indeterminacy. Unlike the first argument, it does not involve reflection on the methodology of radical translation. Instead, it involves reasoning more generally about the nature of theory, language, and meaning. In this argument, Quine claims that indeterminacy of translation results from an underdetermination of physical theory by the evidence. It is important to distinguish this argument from the claim that indeterminacy arises from the underdetermination of theories of meaning in particular; the latter claim is part of the first argument. The strategy of the second argument is immediately puzzling: Why should underdeter-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

mination of theories of physics have anything to do with the scrutability of meanings in natural languages? I turn to the texts here to explicate the argument.

The argument is present in WO, chapter 2, although it is not explicitly developed until later writings. Quine sums up the work in chapter 2, connecting it with the results of chapter 1, as follows:

> Thinking in terms of radical translation of exotic languages has helped make factors vivid, but the main lesson to be derived concerns the empirical slack in our own beliefs. For our own views could be revised into those attributed to the compatriot in the impractical joke imagined; no conflicts with experience could ever supervene, except such as would attend our present sensible views as well. To the same degree that the radical translation of sentences is underdetermined by the totality of dispositions to verbal behavior, our own theories and beliefs in general are underdetermined by the totality of possible sensory evidence time without end (WO, 78).

Quine is claiming here that indeterminacy and underdetermination occur to the same degree, and that the point of exploring the extent of indeterminacy is to show the extent of underdetermination. There is, however, no indication in WO that there is an argument going on here other than the one that takes up most of WO, chapter 2. Later, however, Quine develops the argument more explicitly. In "On the Reasons for the Indeterminacy of Translation," he states that the argument from the methodology of translation is not in fact the real argument for indeterminacy:

> This whole effort [of radical translation] was aimed not at proof but at helping the reader to reconcile the indeterminacy of translation imaginatively with the concrete reality of radical translation. The argument for indeterminacy is another thing, as seen earlier in this paper (p. 182).

The paper begins as follows:

> My gavagai example has figured too centrally in discussion of the indeterminacy of translation. Readers see the example as the ground of the doctrine, and hope by resolving the example to cast doubt on the doctrine. The real ground of the doctrine is different, broader and deeper.
>
> Let us put translation aside for a while and think about physical theory. Naturally it is underdetermined . . . (*ibid.*, p. 178).

Then he goes on to argue:

> . . . the same old empirical slack, the old indeterminacy [sic] between physical theories, recurs in second intension. Insofar as the truth of a physical theory is underdetermined by observables, the translation of the foreigner's physical theory is underdetermined by the translation of his observation sentences (*ibid.*, p. 179).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

So, in this paper, Quine is not only saying that indeterminacy and underdetermination occur to the same degree; he is also claiming that the argument from underdetermination is the "real ground" of the doctrine of indeterminacy and a different argument from the one used in discussion of the gavagai example.

What is the connection between underdetermination of physical theory and indeterminacy of translation? Underdetermination of physical theory is not sufficient for indeterminacy of translation. Underdetermination is a doctrine about scientific theories, not natural languages. Natural languages differ from scientific theories: they are not hypothesized, nor is it clear to what extent they are revisable (see below for more on this). Furthermore, various theories, and nonindicative statements, may be expressed within one natural language. So, underdetermination of natural language does not follow from underdetermination of physical theory. Even if underdetermination of natural language did follow, indeterminacy of translation might not, for indeterminacy is a doctrine about the inscrutability of conceptual differences, not just about the existence of conceptual differences.

Two premises are needed to argue from underdetermination of physical theory to indeterminacy of translation. In fact, if these premises are granted, arguing for underdetermination and indeterminacy becomes the same project. The first premise is that it is appropriate to treat natural languages as scientific theories, at least to the extent that issues of evidence and underdetermination can be raised. The second premise is that the observation sentences, which are those sentences directly keyed to publically observable events, are all that can be determinately translated. These premises yield the result that there is no fact of the matter as to which of various empirically equivalent (having the same observational content) scientific theories the native is expressing. In other words, the extent of indeterminacy is the same as the extent of underdetermination of a scientific theory by its evidence.[30] When Quine writes that indeterminacy of translation arises when "the same old empirical slack, the old indeterminacy [sic] between physical theories, recurs in second intension,"[31] he does not mean, as many have interpreted him, that indeterminacy is the underdetermination of a theory of meaning. Rather, he should be understood literally as saying that the very underdetermination of the most general physical theories recurs at

[30] Considerations of simplicity of theories are not introduced here by Quine. The reason for this will be suggested later.
[31] "On the Reasons for the Indeterminacy of Translation," p. 179.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the level of talk of meanings, and there it is called indeterminacy of translation.

What support is there for the two premises involved in reasoning from underdetermination to indeterminacy? They are not argued for in the texts. Quine talks of translating the native's language and translating the native's theory interchangeably.[32] As I mentioned above, there are important respects in which natural languages are not like scientific theories.

The second premise is even less plausible than the first. One might think that the detailed reflections on the scenario of radical translation—what I have called Quine's first argument—provide the support for the second premise. But there are two problems with reading Quine in this way. First, reflection on the methodology of radical translation does not lead to the conclusion that only the observation sentences can be determinately translated. It almost always[33] leads to the conclusion that translation of truth-functional logic is determinate also, and that the principles of charity and simplicity are objective constraints on translation. So, the arguments are in disagreement on the extent of indeterminacy. Secondly, Quine describes the second argument as the "real ground" and the "deep" set of considerations for indeterminacy which is not affected by any difficulties with the first argument. The two arguments are really different.

There are several passages in which Quine takes empirical meaning to be the whole of meaning, without offering an argument. If empirical meaning is the whole of meaning, then it follows that only the observation sentences will be determinately translatable, since only they have determinate empirical content. Theoretical sentences gain their content through connection with observation sentences, but only as a network of interconnected theoretical claims. So, theoretical sentences have no empirical content that would distinguish one from another, and, in that sense, no determinate empirical content. For this argument, the locus of meaning is empirical meaning, not because of considerations regarding the data and evidence for attributing meaning, but because of a general view about what it is to be meaningful. And the general view is that the facts of meaning are exhausted by the empirical content of the theory that is natural language.[34] In section III, I shall suggest why Quine holds such a

[32] See, for example, the quotations from "On the Reasons for Indeterminacy of Translation" above, and from "Epistemology Naturalized" below.

[33] An exception is the occasion when Quine allows that all meaning is indeterminate. See fn. 17.

[34] This leads to the paradoxical result that all theories with the same content "mean" the same, and in particular that there is no difference between trivial

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

view. Here I show some passages where Quine takes empirical meaning to be meaning. In the following passage from "Epistemology Naturalized,"[35] Quine also treats natural languages as scientific theories:

> We might better speak . . . simply of observational evidence for theories, and we may, following Peirce, still fairly call this the empirical meaning of the theories.
>
> These considerations raise a philosophical question even about ordinary unphilosophical translation, such as from English to Arunta or Chinese. For if the English sentences of a theory have their meaning only together as a body, then we can justify their translation into Arunta only together as a body. There will be no justification for pairing off the component English sentences with component Arunta sentences, except as these correlations make the translation of the theory as a whole come out right. Any translations of the English sentences into Arunta sentences will be as correct as any other, so long as the net empirical implications of the theory as a whole are preserved in translation. But it is to be expected that many different ways of translating the component sentences, essentially different individually, would deliver the same empirical implications for the theory as a whole; deviations in the translation of one component sentence could be compensated for in the translation of another component sentence. Insofar, there can be no ground for saying which of two glaring unlike translations of individual sentences is right (*ibid.*, pp. 79–80).

In this passage, there is a slide from 'empirical meaning' to 'meaning'[36], as well as a slide from 'English sentences of a theory' to 'English sentences' and 'ordinary translation'.

---

variants of the same theory and real theoretical alternatives. Quine has to claim that the differences between theories are merely terminological, and apparent conceptual differences the result of our confusion. The difficulty is that he does not explain why some theories seem conceptually closer than others. Quine has worried about this recently [see the early essays in *Theories and Things* (Cambridge: Harvard, 1981); and "Relativism and Absolutism," *Monist*, LXVII (1984), pp. 293–296] but has come to no satisfactory conclusion.

[35] In *Ontological Relativity*.

[36] This interpretation of the passage deserves some more substantiation. For it occurs within a section of the paper which Quine begins by saying: "suppose we hold, with the old empiricist Pierce, that the very meaning of a statement consists in the truth it would make to possible experience" (*Ontological Relativity*, p. 78). So, one might think that Quine is not adopting this empiricism at all, but, rather, just examining its consequences. I do not take this interpretation, for several reasons. Quine's tone is not one of simple hypothetical argument. He takes on the assumption in the context of a general attraction toward empiricism and draws out consequences—notably, the indeterminacy of translation—that he endorses. And although he concludes the two pages of reasoning under supposition as follows—"Should the unwelcomeness of the conclusion persuade us to abandon

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The suggestion that empirical meaning comprises all the facts of meaning can also be found in WO. At the beginning of chapter 2, Quine writes that we must look to "non-verbal stimulation" for "whatever empirical content there may be," and that his aim in the chapter is to see how much "empirically unconditioned variation" in conceptual schemes there is (p. 26). This is not the claim that we can only use verbal behavior as evidence for translation (which is the first argument for indeterminacy). One difference is that behavior in response to *verbal* stimulation is not included as part of the objective basis for translation in this passage at the beginning of chapter 2 of WO. Another difference is that talk of nonverbal stimulation as the locus of meaning is quite different from talk of behavior as the data for translation. For the latter discussion, considerations of simplicity and charity may enter in deciding on a translation, but not for the former.

So, the second assumption in the second argument for indeterminacy rests on the view that empirical meaning is all the meaning there is. On this view, the facts of meaning are a *particular* set of behavioral facts—namely, the facts concerning the conditioning of sentences to nonverbal stimulation. Despite the use of the word 'behavioral', it is more an empiricist assumption than a behaviorist one, since it involves the view that empirical content is all of meaning. The assumption is couched in terms of behavior, since language gets its empirical content through behavioral conditioning to nonverbal stimulation (rather than, say, by "naming sense data" or some such traditional empiricist device).

A recent paper, "Indeterminacy of Translation Again" also shows both arguments for indeterminacy. Quine has the first argument in mind when he reasons that "[e]ach of us learns his language by observing other people's verbal behavior. . . . We depend strictly on overt behavior in observable situations" (p. 5). That is, the *data* for language learning are limited to behavioral facts about the speakers

---

the verification theory of meaning? Certainly not. The sort of meaning that is basic to translation and to the learning of one's own language is necessarily empirical meaning and no more" (*Ontological Relativity*, p. 81) [Quine goes on to allude to the detailed considerations for indeterminacy which constitute the first argument] —this does not alleviate the sense that he is doing more than just supposing that empirical meaning is all the meaning there is. For Quine writes as though the supposition has antecedent plausibility, but suggesting that, in the face of unwelcome conclusions (viz., indeterminacy) perhaps we should abandon it. Indeed, who said that we should adopt it in the first place?

Furthermore, it should be pointed out that the detailed considerations that Quine gives do not in fact support his conclusion that empirical meaning is all the meaning there is. They either lead to complete skepticism, or, usually, to the conclusions that more than the observation sentences can be determinately translated. Clearly, the two arguments for indeterminacy are less compatible than Quine makes out.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

of a language. He has the second argument in mind, however, when he writes, "there is only the native's verbal behavior for the manuals of translation to be right or wrong about" (p. 9), for here he is taking the *facts* of meaning to be exhaused by the facts of behavior.[37] This explains Quine's puzzling statement at the beginning of the paper, "[i]n psychology one may or may not be a behaviorist, but in linguistics one has no choice" (p. 9). Quine's behaviorism here amounts to a specific view regarding the facts of linguistic meaning. If his behaviorism was merely a call to base psychological attributions on observable data, then linguistics would be treated no differently from other topics in psychology.

The two arguments for indeterminacy are not compatible with one another, since they reach different conclusions on the extent of determinate translation. The arguments are even opposed in spirit; according to the second argument, the point of indeterminacy is to show the extent of possible variation in conceptual scheme. But the first argument, which often makes use of the principle of charity, attributes to the native the beliefs that *we* take to be the most obviously true. And, on the occasion that the first argument goes completely skeptical, with the result that none of translation is determinate, the second argument, which appeals to the determinate nature of empirical content, is blocked. The two arguments for indeterminacy involve different kinds of reasoning. The first is a skeptical argument based on the data and methods of the ordinary practice of translation. The second involves general considerations about what the facts of meaning are and crucially employs two assumptions: that natural languages are like scientific theories and that empirical meaning is all the meaning there is.

The two kinds of reasoning employed in the arguments for indeterminacy also show in Quine's treatment of other topics. He either considers actual features of language and scientific practice and is mostly sympathetic to ordinary standards of objectivity, or he invokes general assumptions about the nature of language and empirical significance. This is evident in his varying remarks on the status of canons of scientific method such as simplicity and conservatism,[38] his

[37] This is not, of course, the second argument for indeterminacy in full. The second argument involves more than the claim that the facts of meaning are the behavioral facts. It involves the claim that the facts of meaning are a *specific set* of behavioral facts. I think enough of the second argument is present, however, to justify reading the text in this way.

[38] Sometimes Quine holds that simplicity is in itself an objective criterion of theory choice, so that, even in the limit of all possible evidence, the simpler theory is objectively preferable (e.g., "On Empirically Equivalent Systems of the World," *Erkenntnis*, IX (1975), pp. 327/8). Sometimes, however, he says the reverse (e.g., "Replies to Eleven Essays," *Philosophical Topics*, XII (1981), pp. 227/8).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

recent vacillation on the possibility of alternative theories in the limit of all possible evidence,[39] his difficulty with, but attempts to make sense of, an ordinary notion of theory,[40] and his occasional taste for dispensing with theory in favor of merely the observational content of the theory.[41] On the one hand, Quine is sympathetic to ordinary insights such as regarding simplicity (whatever that is) as part of scientific method, drawing a distinction between variants of the same theory and different theories, and holding that only one distinct theory of the world can be true. On the other hand, Quine employs some kind of empiricist assumptions: regarding simplicity as not objective when it does not contribute to predictive success, thinking that underdetermination in the limit of all possible evidence is inevitable, and regarding theory that goes beyond empirical content as here on sufferance.

One might be tempted to end here and conclude that there are two different strains of thinking in Quine's work. Such a conclusion would leave much unexplained. The persistence of two *contrary* strains of thinking invites explanation, as does Quine's statement that the second argument for indeterminacy is the "deep" and "real" ground of the doctrine. The nature and source of the empiricist assumptions behind the second argument have not yet been explored. And we do not understand Quine's conviction that there is indeterminacy, a conviction which persists through his vacillations on what the argument for indeterminacy is.

III. NATURAL EMPIRICISM

At the heart of Quine's work is the following unargued view. Persons are physical beings whose cognitive and linguistic lives consist in learned verbal responses to sensory stimulation. These verbal dispositions are learned for one end: to predict successfully (and hence control) sensory, nonverbal stimulation. The dispositional system is at the same time the person's theory of the world and the person's

[39] See the end of "On Empirically Equivalent Systems of the World," where he claims that alternative theories are possible; and "Things and Their Place in Theories" (*Theories and Things*, pp. 29–30), where he claims that they are not possible. In the latter paper, he makes use of a logical device to escape the claim that there is underdetermination. But such a device is in place only to serve the ordinary view that just one theory of the world can be true. I argue this at length elsewhere. See the beginning of fn. 20.

[40] Compare "On Empirically Equivalent Systems of the World," p. 319; "Empirical Content" (*Theories and Things*, pp. 24–30); "On The Very Idea of a Third Dogma" (*Theories and Things*, pp. 41/2); and "Relativism and Absolutism," pp. 293–296). In these papers, a tension shows between Quine's desire to support the ordinary distinction between variants of the same theory and different theories (he rarely abandons the distinction), and his view, deriving from the indeterminacy thesis, that making such a distinction is not possible.

[41] This shows in the main body of "On Empirically Equivalent Systems of the World."

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

language. In other words, the generation of language is the generation of science, and science aims to maximize predictions. Thus, it is reasoned, the observational content of a language is its significant content, and language gets its observational content through learned dispositions to sensory, nonverbal stimulation. Since language is so extensively tied in with human functioning, the view may be summed up by saying that Quine regards people as walking theories[42] and theories as devices for generating predictions.

I call this view *natural empiricism* for two reasons. First, the claim is that in the natural state—not just when doing science—a person aims to generate predictions. In other words, we are naturally empiricists. Secondly, Quine espouses a particular form of empiricism, to be distinguished from the phenomenalistic empiricism of, for example, Berkeley, Ernst Mach, and early Rudolf Carnap. Quine's view is that the empirical content of language consists in dispositions to respond to sensory, nonverbal stimulation. There is no appeal to internal experience, awareness, or sensation, and certainly no attempt to construct the world out of such experience. Language is connected to sensory stimulation by dispositions, rather than by any intentional acts of naming. Quine's views are completely and immediately naturalistic.

There are passages in the texts, often at the beginning of papers, in which Quine presents parts of natural empiricism. They are presented in a tone of stating the obvious, without argument. Examples of these passages are:

> As an empiricist I continue to think of the conceptual scheme of science as a tool, ultimately, for predicting future experience in the light of past experience.[43]

> I am a physical object sitting in a physical world. Some of the forces of the physical world impinge on my surface . . . I strike back, emanating concentric air waves . . . All I am or ever hope to be is due to irritation on my surface, together with such latent tendencies to respond as may have been present in my original germ plasm.[44]

> We have been reflecting in a general way on how surface irritations generate, through language, one's knowledge of the world. One is taught so to associate words with words and other stimulations that there emerges something recognizable as talk of things, not to be distinguished from truth about the world. The voluminous and intricately structured talk bears little evident correspondence to the past and present barrage

[42] I am grateful to Hilary Putnam for this phrase expressing my views.
[43] "Two Dogmas of Empiricism," in *From a Logical Point of View* (New York: Harper & Row, 1953, 1961), p. 44.
[44] "The Scope and Language of Science," in *Ways of Paradox* (Cambridge: Harvard, 1976), p. 228.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

of non-verbal stimulation; yet it is to such stimulation that we must look for whatever empirical content there may be (WO, 26).

Our talk of external things . . . is just a conceptual apparatus that helps us to forsee and control the triggering of our sensory receptors in the light of previous triggering of our sensory receptors.[45]

These passages involve more than the claim that we should be studied as physical objects, or the claim that behavioral explanations of the mind are best. They involve the further views that there is an aim of knowledge, viz., predictions, that "all we are," and, in particular, all our linguistic capability, is due to dispositional mechanisms keyed to surface irritations (with the exception of a few innate dispositions), and that the content of language is due to the nonverbal surface irritations. It is these further views that distinguish natural empiricism from the lesser commitments of physicalism and behaviorism.

So, there is textual evidence that Quine holds the views that constitute natural empiricism. Now, I show that it is these views—rather than simply behaviorism or physicalism, for example—which generate his arguments for indeterminacy and other central doctrines. It is only when I have done this that it is appropriate to conclude that natural empiricism is Quine's central view.

The second argument for indeterminacy is a straightforward consequence of natural empiricism. Indeed, one might say that indeterminacy of translation is a dramatic presentation of natural empiricism. The first assumption—that natural languages are like scientific theories of the world—is part of natural empiricism, which holds that both natural language and science are constrained only by their aim of successfully generating predictions. Thus, they are similarly underdetermined by the evidence. The second assumption—that empirical meaning is all the meaning there is—results from natural empiricism if one reasons as follows. The aim of all theory is predictive power. Hence, all that matters to a theory is its empirical content, and the precise mechanism of incorporating that content, i.e., the theoretical superstructure, is arbitrary. Using the first assumption, all that is significant to the theory that is language is empirical content. Hence, empirical meaning is all the *real* meaning there is.

This close connection of the second argument for indeterminacy with natural empiricism explains several things. It explains why Quine considers the second argument to be the "deeper" argument, for the second argument flows directly from Quine's deepest views, whereas, as we shall see below, the first argument does not follow as

[45] "Things and Their Place in Theories," in *Theories and Things*, p. 1.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

immediately. It explains Quine's conviction that there is indeterminacy, despite the difficulties with the first argument—given his conviction that natural empiricism is true. One could also turn this around and say that, for Quine, there *has* to be indeterminacy, or else the underlying picture of human nature, i.e., natural empiricism, is wrong. For suppose that our ordinary practices of translation were objective, and thus there were no indeterminacy. We see little conceptual variation between languages, and this tells against the view, central to natural empiricism, that language is a theory constrained only by the goal of predictive success. This explains why indeterminacy of translation is so central and so firm a doctrine for Quine.

Natural empiricism also generates the first argument for indeterminacy, although less immediately than the second argument. It does so by first generating the doctrine of "working from within" as follows. According to natural empiricism, speaking a language involves the use of an evolving system of verbal dispositions conditioned to stimulation. Some of language—for example, logic, mathematics, and many general and theoretical claims—is quite abstract, not having isolatable empirical content. This tells us that the system of verbal dispositions is a complex network in which some parts—the parts with no isolatable empirical content—interconnect with many others. It follows that alteration of verbal dispositions will often be a complex matter, resulting in widespread linguistic change. It is also the case, according to natural empiricism, that, insofar as we talk, we have linguistic dispositions. So, in describing and evaluating anything, we can at first only do so from within the verbal dispositions that we have. These can change, but it is not expected that change of our most general terms and general views can take place locally. Rather, it is expected that change at the theoretical level will involve readjustments throughout the system of verbal dispositions. If these readjustments are not made, general terms break off from their connections with other verbal dispositions and hence lose empirical content, becoming meaningless. So, the attitude of working from within ordinary language and theory, starting with what we have and shifting conceptual systems only gradually and with care, is part and parcel of natural empiricism.

Quine's first argument for indeterminacy is the result of working from within the ordinary conceptual scheme. His expectation from natural empiricism is that indeterminacy will show from within the ordinary practice of field linguists. If it does not, the meaningfulness of the standards of objectivity used in the second argument is doubtful. So, Quine immerses himself in the practice of radical translation and asks how much of the method is objective. He claims

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

that subjective maxims are being used in translation—projection of our own conceptual scheme and our parochial standards of simplicity, for example. These are maxims which we ordinarily regard as subjective. He also claims that some translational practices result in objectively better translation manuals—for example, use of the maxim of (objective) simplicity and the principle of charity. These are maxims employed by field linguists. Although an explication of them would be of interest, they are currently regarded as yielding objectively preferable translation manuals. So, the first argument results in including constraints such as simplicity and the principle of charity as objective constraints on translation.[46]

The two approaches that Quine takes to issues of underdetermination and scientific method, which I mentioned briefly in section II, may be understood as arising from natural empiricism in the same way as do the two arguments for indeterminacy. Natural empiricism leads to the view that only the empirical content of a theory matters; even simplicity (where it does not contribute to empirical content) is not objective, since it is not traceable to the world that is being described by the theory. If, as is widely held, there is more than one theory that captures all possible evidence, then more than one theory will be acceptable in the limit of all possible evidence. And if, with natural empiricism, it is held that only the empirical content of a theory matters, it is a short step to go further and try to dispense with what does not matter to theory, viz., theoretical material. It also becomes impossible to claim that there is a difference between theory variants and deeply different theories. Theories can only differ terminologically. So, these general views on underdetermination and scientific method may be understood directly in terms of natural empiricism. Also, Quine's sympathy to ordinary insights regarding scientific method and truth may be understood as generated indirectly from natural empiricism, through the doctrine of "working from within."

Thus, there is a tension in Quine's work between the direct consequences of natural empiricism and the consequences of "working from within." This is the major tension in his views. Many philosophical works include both ordinary and theoretical strains that are in

[46] I argued above that the first argument is overly skeptical, and thus not an argument from within the ordinary practice of translation. Indeed, I pointed out how the skeptical style of argument, if used consistently, leads to a result that Quine is unwilling to accept—indeterminacy throughout. When I say here that the first argument is an argument "from within," I am claiming that Quine is trying, from within, to show that translation is indeterminate, and that he is somewhat sensitive to ordinary objective maxims on translation such as simplicity and the principle of charity. To make his argument work, he does become overly skeptical, and this makes the argument unstable in the ways that I have pointed out.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

conflict with one another. But Quine's views are unusual in that the conflict is generated internally, since the doctrine of "working from within" can be derived from natural empiricism. So, the tension is there to stay, and thus it is a problem for his views.

IV. FURTHER CONSEQUENCES, EVALUATION AND CONCLUSIONS

The doctrine of underdetermination, which leads to indeterminacy, is a central consequence of natural empiricism. It involves more than the claim that several theories have the same empirical consequences. It also involves the claim that several such theories can be held by us humans. Only after this is established can Quine go on to argue that there is no fact of the matter regarding *which* theory a person has adopted. The claim that there is underdetermination is a *psychological* claim that we *can,* and in fact *do,* hold a variety of theories generating the predictions. So, it is denied that we might be seriously constrained[47] in our choice of theory by prexisting structures and dispositions, or by inextinguishable beliefs.

This psychological claim results from natural empiricism as follows. Natural empiricism is the view that we are psychologically structured to learn dispositions to stimulation so as to maximize predictive power. The psychological, or natural, aim is the rational aim. If there were other psychological constraints on theory choice, then we would be psychologically structured to do more than pick the theory in conformity with observations. Then Quine would have two choices, both of which go against natural empiricism. Either he can say that the rational aim is more than generating predictions, and continue to identify the natural with the rational. This would involve giving up the claim to a purely empiricist account of rationality. Or Quine can drive a wedge between the natural and the rational and allow that the rational aim—to maximize predictions—is not the natural aim, and, indeed, our natural activity might even frustrate that aim. This is an unwelcome result, since it involves giving up one of the characteristic features of natural empiricism, namely, the claim that rationality is grounded in our natural activity. This result would also involve giving up the optimism of natural empiricism, which comes from the idea that one can adapt dispositionally to any set of stimulations. With more rigid psychological structures, there is no longer that guarantee. So, it is easy to see why Quine is committed to the doctrine of underdetermination. What underdetermination means for Quine is that theory choice is not constrained by psychological states and structures (i.e., by constraints that he does not want

[47] A small degree of prior constraint on dispositions to be developed is tolerable and even necessary (see Quine's discussion of similarity standards in "Natural Kinds," in *Ontological Relativity*). What I say here and subsequently refers to serious and, especially, unrevisable internal constraint.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

to regard as rational). If the only rational constraint on theory choice is conformity to observation, it is unavoidable that there be some theoretical material not chosen because it has greater predictive power than other adequate theories, but chosen in some other way. Quine's suggestion is that theory choice is random, which makes the theoretical commitments less insidious.[48]

Now that natural empiricism and its consequences have been described, it is time to make an assessment. Hume notwithstanding, Quine is the first since the scientific revolution to put forward a thoroughly naturalistic and thoroughly empiricistic world view. In doing so, he emphasizes two central scientific values: the importance of predictive success as a criterion of choice between scientific theories, and the materialism of current scientific theory. Experience serves to ground all of knowledge and language: to get us true theories and meaningful utterances. And the empiricism emerges from, and remains in, a view of a person as a physical being whose only input from the world is the stimulation of nerve ends. His empiricism is optimistic, carrying with it the expectation that we can intellectually adapt to whatever we might find, merely by relearning dispositions. It is not surprising that his views attract so many with empiricistic and scientistic inclinations. These attractive features are not sufficient to support natural empiricism, however. The view has a number of difficulties, to which I turn now.

According to natural empiricism, natural languages are like scientific theories. As mentioned in section II, this is questionable. Natural languages are not hypothesized, and there is reason to think with Chomsky[49] and Fodor[50] that there are limits on the revisability of the

[48] I argued in section III that the second argument for indeterminacy makes use of the doctrine of underdetermination, which comes from natural empiricism. There is a troubling consequence of indeterminacy, however, which I did not note: according to indeterminacy, allegiance to a theory is not a matter of having a particular set of dispositions to stimulation, for there is no fact of the matter which theory one has, whereas there is a fact of the matter which set of dispositions one has. So, we could have a fixed, even innate, set of dispositions to stimulation and it still be indeterminate which theory we hold. So, in the end, indeterminacy is inconsistent with the view that holding a particular theory is having a particular set of dispositions to stimulation. This inconsistency is due, in part, to the assumption from natural empiricism that empirical meaning is all the meaning there is. It is indicative of the internal difficulties with natural empiricism, to be commented on shortly.

[49] For example, see *Rules and Representations* (New York: Columbia, 1980), p. 46, where Chomsky writes that "there is an inseparable connection between the scope and limits of human knowledge." He spells out the claim in the course of ch. 1.

[50] For example, see *The Language of Thought* (New York: Crowell, 1975), pp. 96/7, where Fodor writes of the limits of one's conceptual system; and also see *The Modularity of Mind* (Cambridge: MIT, 1983), pt. v, where Fodor shows "modified rapture" (p. 129) over his conclusion that our thought processes are partly modu-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

structure of natural languages. Furthermore, alternative theories and nonindicative statements may be expressed within one natural language.

A central claim of natural empiricism is that the only aim of theory (which includes both scientific theory and natural language) is predictive success. It is likely that predictive success is the most important test of scientific theories. But it does not follow from this that predictive success is their aim. Also, other constraints on theory choice, such as simplicity, are operative. Furthermore, our ordinary conceptualization of the world seems not to be aimed merely at accurately predicting sensory stimulation (if indeed it is aimed at anything). And, if Quine's claim rests on the view that what we *naturally* do is aim to generate predictions, it must be replied that we naturally do many other things: indeed, we devise theories of the world which satisfy emotional, ethical, or political needs. Just as 'natural' is too weak to carry 'ethical', it is too weak to carry 'rational'.

In drawing out the important consequences of natural empiricism —underdetermination and indeterminacy—doubtful moves are made. First, as explained at the beginning of this section, Quine's claim that there is underdetermination is a psychological thesis, susceptible of empirical test. Quine does not suggest putting his views to such a test, however. There is accumulating evidence that we are predisposed to learn particular linguistic structures.[51] This evidence tells against Quine's views. Secondly, in the second argument for indeterminacy (as seen in sections II and III), Quine moves from saying that predictive success is the only aim of theory to saying that empirical meaning is the only meaning there is. In this move, he slides from saying that empirical content is all that *matters* to language to saying that there is no other content of language. Although this slide is not unmotivated, it confuses what is significant with what is a fact of the matter.

---

lar. He claims that it is an empirical question how much of thought is modular, and how much may be understood on an "isotropic," "equipotential," or "Quinean" model. He also cautions against regarding the isotropic model as an epistemically unbounded model of mind, since, even on an isotropic model, the units of thought are fixed. It is interesting to note in this connection that Quine does seem to be attracted (perhaps wrongly) to the idea of epistemic unboundedness, and also that Quine's model of mind is less epistemically bounded than is Hume's (the example Fodor gives), for natural empiricism allows conditioning of sentences to any detectable state of affairs (concepts are indeterminate). There is still some boundedness here, imposed by the limitations of our sensory equipment, but it is less than the boundedness in Hume's work, which comes from the fixed vocabulary of ideas. Quine is a more thoroughgoing empiricist than Hume.

[51] See fn. 49 and fn. 50.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Finally, as discussed in section III, there is a tension in Quine's work between the results of "working from within" and the direct consequences of natural empiricism. I also argued that "working from within" follows from natural empiricism, so the tension is internal. Hence, there is serious internal inconsistency in Quine's work.

These difficulties render natural empiricism doubtful as a whole. Natural empiricism is not doubtful, however, in the way that a discredited scientific theory is. Although natural empiricism has some testable parts, it is more characteristic of a metaphysical than a scientific theory.[52] For the view is held to with conviction, and often repeated, without argument, and without effort to put it to empirical test where appropriate. Furthermore, as noted above, not all of the reasoning from natural empiricism is as careful as the reasoning in a scientific theory.

Although natural empiricism may be doubtful as a whole, it does not follow that each part of the view is doubtful, and therefore it does not follow that all of Quine's arguments from natural empiricism may be dismissed. The second argument for indeterminacy rests on the least plausible parts of natural empiricism and is discredited, but the first argument, which starts from "working from within," needs more careful assessment. I indicated in section I that the argument is overly skeptical, and suggested lines to pursue in evaluating it, but I did not complete my cricitisms. Skeptical doubts cannot be fully put to rest until this is done.

Quine's confidence that there is indeterminacy, and his repeated claims that it follows from objective consideration of the data, or from physicalism, or from behaviorism, has, rightly, made philosophers cautious in evaluating his views. I have given an interpretation of Quine's writing which makes sense of all the different strains, and shows a single view to lie behind them. This greater understanding of the arguments frees us to reject some and to continue consideration of others. It turns out that Quine's most central arguments are not the most promising arguments for us. In this sense, his work has been put into historical perspective.

MIRIAM SOLOMON

University of Cincinnati

[52] I do not mean to suggest that there is a clear line between metaphysical and scientific theories.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

COMMENTS AND CRITICISMS

JOHNSTON ON HUMAN BEINGS*

MARK Johnston[1] has recently proposed a solution to the problem of personal identity, in which persons are claimed to be essentially human beings, a kind distinct from the merely biological entities that are human organisms.

After criticizing wide psychological reductionism coupled with the method of cases and proposing that we begin our theorizing with the "humble and ubiquitous practice of reidentifying each other over time" (63), Johnston gives his positive account (which I shall discuss here), which he rests on Shoemaker's case of Brown and Robinson, where Robinson receives Brown's brain in a transplant. The predominant reaction to the case, says Johnston, is that Brown is Brownson (namely, the person with Robinson's body and Brown's brain). Thus, although the alternative method will begin by reidentifying persons as human organisms, i.e., biological entities, it seems we have to depart from this approach, since, in bizarre cases such as this one, the brainless body of Brown (which Johnston calls "Brownless") is the same organism that Brown was, but Brown is now Brownson. The reason for this reaction to the case (which Johnston regards as a correct reaction) is, quite simply, a Lockean one: Brownson's tracing of himself backward in time relies on memory; memory claims by which one traces oneself essentially involve claims about personal, not merely mental, identity; Brownson's memory claims are *ex hypothesi* the same as Brown's; therefore, Brownson is Brown.

Before looking at Johnston's own characterization of persons, it must be remarked that it is not at all clear that our predominant reaction to Shoemaker's case is to suppose that Brownson is Brown and what this supposition implies, namely that Brown survives during the transplant as his disembodied brain. *If* this is our reaction and the reaction is justified, then Johnston may be right to claim that persons are not essentially human organisms; but it is quite conceivable that with extra information about brain-stem material in Brownless we do *not* have the intuition that Brownson is Brown, even though Brownson is psychologically continuous with the preoperative Brown. Our intuitive reaction to Brownless might be the same as it would be to Brown after he suffered massive brain damage in an

* I would like to thank David Pears and Derek Parfit for their helpful comments on an earlier draft of this paper.

[1] This JOURNAL, LXXXIV, 2 (February 1987):59–83.

0022-362X/89/8603/137–141 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

accident, or during the terminal stages of a degenerative brain disease: of course, we might judge that the person before us is not the same person as Brown, qualitatively speaking, in which case we would, more properly, say that Brown is not the same person as he was, but we would still affirm numerical identity.[2] And the case raises the sorts of questions Derek Parfit discusses concerning the psychological spectrum (*ibid.*, pp. 231–233): just how much of Brown's brain must be removed before we no longer affirm that he is the organism before us? If his brain were removed in parts over a period of time, at what point would we deny numerical identity of the person? It seems that our intuitions here are by no means as clear as Johnston seems to think they are.

Now, the fact that Brownson is, *ex hypothesi*, fully psychologically continuous with Brown, cannot be relied on by Johnston to justify by itself the intuition that Brownson is Brown, since that would be to take the reductionist point, and it would then be a short step to acknowledge that *identity* did not matter. Instead, what Johnston relies on (which, together with Brownson's memory claims, yields the conclusion that Brownson is Brown) is the fact that it is Brown's own brain which has been transplanted into Robinson's head. Hence his proposal that persons are human beings which, although normally reidentified in terms of human organisms, are not the same as human organisms, since a human being can outlive the organism to which it belongs (in some sense of 'belongs'). We are then asked to believe that the person Brown, during the transplant, survives as a "radically mutilated human being," to wit a brain. Now, this two-fold characterization of human beings and human organisms is both ad hoc and obscure. Just what are these human beings? Johnston himself acknowledges his rather stipulative use of the term 'human being' (64). He certainly cannot mean what we ordinarily mean by 'human being', since what we ordinarily mean, if indeed there is any definite meaning at all, is "human organism," "member of the species *homo sapiens*." Just what Johnston's human beings are is quite unclear: a debodied brain is claimed to be a radically mutilated human being, though not a radically mutilated human organism. Presumably then, when it was embodied, the human being was not radically mutilated. Neither was the organism. So a healthy, unmutilated member of *homo sapiens* is both a being and an organism [though the being is not a separately existing entity—Johnston wants to deny this (63, fn. 7)], and upon mutilation can be one or the other without being both. (If you take away everything but its brain, what

[2] See Parfit, *Reasons and Persons* (New York: Oxford, 1984), pp. 201/2.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

you have left is a being without the organism; if you take away its brain, what you have left is the organism without the being.) It seems that what Johnston is getting at is that the "beingness" of the human organism is a function of its capacity to support a (complex) mental life. But we have *no* knowledge (and may never have knowledge) of the capacity of a disembodied brain to support *any* mental life: after all, *it is only when embodied that the brain demonstrates any mentality at all.* It is invalid to argue, from the fact that a human organism with a severely damaged brain has an impoverished mental life, to the fact that a brain *on its own* can support a mental life. In which case Johnston might be forced back to the position that only the whole organism, i.e., at least a body and a brain, is a person. But it is open to the reductionist to reply that, if it is the mental life of the human about which we are concerned, then it is a mere prejudice on Johnston's part to require, as a condition on Brown's survival, that some *physical* part of Brown, i.e., his brain (or, for all we know, his brain and a good deal of the rest of his body), persists through time as the bearer of Brown's mental life: it is enough if Brown's mental life itself is recreated. Why then could Brown not survive teletransportation, whereby an exact mental and physical replica of himself appeared on Mars? If mental life is what matters in personhood, as Johnston seems to think as much as the psychological reductionist, then why stop with the brain and not allow the human being (= person) to be traced, not merely through a brain transplant, but through full teletransportation? That is to say, why does the dispositional and occurrent mentality *itself* not constitute the human being, irrespective of the particular brain tissue in which it is realized?

The intuition generated by Bernard Williams's first presentation of his conundrum (whereby a doctor uses a machine to transfer *A*'s dispositional and occurrent mentality into *B*'s brain and vice versa, producing the same effects as would be produced by brain transplantation) is that *A* and *B* swap bodies when their dispositional and occurrent mentalities are swapped. This reaction is incompatible with Johnston's theory, which generates the result that *A* could only survive in what was *B*'s body if there were a brain transplantation, and so Johnston suggests that our intuitive reaction to Williams's first presentation is the result of distortion by what Johnston calls the "psychological- and social-continuer effects." Given that he is keen to begin his theorizing about personal identity from a consideration of our ordinary practice of reidentification of persons, however, it is hard to see why the psychological- and social-continuer effects are seen as *distorting* influences on the metaphysical truth of the matter. After all, such influences are a central part of our practice of rei-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

dentifying persons, and inform that practice. The inclination to trace *A* through a body swap and so reidentify him in the *B*-body is generated by the knowledge that the person associated with the *B*-body will be a unique psychological and social continuer of the preoperative *A*. Despite Johnston's reliance on reidentification as a "reliable and unproblematic" source of knowledge about personal identity, however, he stipulates that our practice in the case of Williams's conundrum is distorted. Why? Because it does not conform to the truth of the matter, which is that *A* and *B* have not swapped bodies. Why is this the truth of the matter? Because of the stipulation that a condition on survival is that some *physical* part of the person, viz., his brain, persist as the bearer of the person's mental life. The reductionist, however, rejects all of this. Our tendency to trace *A* through a body swap is correct, he says, not merely as a tendency in conformity with the "practical concomitants of personal identity," but as a tendency which reveals the truth that psychological continuity per se is all that matters in survival.

Johnston also claims that, in any case, there is something arbitrary in our inclination to trace *A* through a body swap: "it is difficult to get the uninitiated to come up with the body-swap intuition in response to the first presentation of Williams's case when the first presentation comes immediately after cases that emphasize the importance of bodily continuity" (81). He claims that our intuitions in such a case vary in accordance with the types of case to which the case at hand is assimilated. To explain our intuitive reaction by appealing to a survey is specious. The second presentation of Williams's case, whereby *A* is shown to survive radical psychological discontinuity (a doctor has to perform a painful operation on *A* without anaesthetic; he tells *A* he can bypass the pain by causing radical psychological discontinuity in him, and by reorganizing *A*'s dispositional and occurrent mentality after the operation so that it is continuous with *A*'s preoperative state), is said to stress the importance of bodily continuity. And yet, although presented at the same time as the first presentation, it does not seem to alter our intuition in the latter that *A* and *B* have swapped bodies. On the contrary, the whole point of the example for Johnston is that our intuitions in the two presentations stand in stark contrast!

In conclusion, apart from the apparent doctrinal difference between Johnston and supporters of reductionism (i.e., concerning teletransportation and, generally, the cut-off point for survival, which Johnston draws, unjustifiably, I claim, at the boundary of persistence and nonpersistence of a certain type of proper physical part of the person, and which Parfit draws at the boundary between

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the persistence and nonpersistence of a chain of continuous mental life irrespective of the *particular* physical matter in which it is realized) it is hard to see where else any differences may lie. For it appears that, in the puzzle case of fission, where *A*'s brain is split and transplanted into the skulls of *B* and *C*, the deliverances of both Johnston's and the reductionist's theory will be the same, viz., that *A* survives as *B* and *C* though is neither one of them. What the reductionist will then point out, however, is that identity cannot be what matters in survival, for the familiar reasons. And if Johnston accepts this, as arguably he should, then his theory will not be one of personal *identity:* it will simply be a weaker form of reductionism.

DAVID S. ODERBERG

Wolfson College/Oxford

BOOK REVIEWS

Collected Papers. GARETH EVANS. New York: Oxford University Press, 1985. 407 p. $39.95.*

Gareth Evans's *Collected Papers* contains all of his published papers. A little over half the volume deals with reference and related topics in semantics and the philosophy of language. These are the articles with which I shall be concerned.[1]

In my judgment, Evans's best work in semantics is his discussion of pronouns in articles 4, 5, and 8. Two of his negative theses stand out as having lasting significance. The first is that some occurrences of pronouns with quantified antecedents, such as those in (1), do not function either as variables bound by their antecedents, or as pronouns of laziness (which are equivalent to repetitions of the antecedents).

(1a) *Few students* came to the party, but *they* had a good time.
(1b) If Mary buys *a present,* then *it* will be expensive.

* I thank Gilbert Harman, Saul Kripke, David Lewis, Stephen Neale, Lorenz Puntel, Nathan Salmon, and George Wilson for their comments.

[1] One article, "The Causal Theory of Names," is well enough known not to require extended comment. Although originally seen by Evans as a way of reconciling description theories with Saul Kripke's insights in *Naming and Necessity* (Cambridge: Harvard, 1980), it has largely been superseded by Evans's more recent *The Varieties of Reference* (New York: Oxford, 1982). In retrospect, its most salient points seem to be its "Madagascar" example of reference change and its emphasis on the role of speakers' intentions in determining and transmitting reference.

0022-362X/89/8603/141–156 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The second negative thesis is that the contrast between the examples in (2) and those in (3) is not the result of a rule that requires the noun/pronoun pairs in (2) to differ in (intended) reference, but rather is the result of a rule that allows the pronouns in (3) to derive their interpretations from their antecedents.[2]

(2a) He loves John's mother.
(2b) He is happy when Bill is in love.
(3a) *John* loves *his* mother.
(3b) *Bill* is happy when *he* is in love.

One of Evans's key positive insights is that the pronouns in (3) have a reading in which they function in the same way as those in (4).

(4a) *Every man* loves *his* mother.
(4b) *Every man* is happy when *he* is in love.

Evans uses this insight to argue that the (bound) pronouns in (3) are not ordinary demonstratives that refer to objects whose contextual salience has been raised by being mentioned earlier in the sentence. Since the role of the antecedents in (4) is not that of raising contextual salience, the role of those in (3) is not that either.

How then should (3) and (4) be unified? According to Evans, the rule interpreting (3) stipulates simply that an occurrence of a pronoun with a singular term antecedent refers (rigidly) to the same thing as its antecedent. A quantified sentence ⌜Every F is G⌝ is said to be true iff ⌜N is G⌝ is always true when N is a (new) name that refers to anything in the extension of F. Pronouns with quantified antecedents are interpreted by applying the rule for quantifiers followed by the rule for pronouns. Thus, occurrences of pronouns in (4) are assimilated to those in (3), even though, according to Evans, those in (3) refer to particular individuals, whereas those in (4) do not.

Despite Evans's evident insights, there are potential problems with this treatment of bound pronouns. First, it depends on the contro-

[2] The target of Evans's first negative thesis is Peter Geach, *Reference and Generality* (Ithaca: Cornell, 1962); the target of the second is Howard Lasnik, "Remarks on Coreference," *Linguistic Analysis*, II (1976): 1–22. Lasnik's views, which have been influential in generative grammar, are based on the observation that demonstrative uses of pronouns can refer to any conversationally salient object. Since one way of raising the salience of an object is to mention it, Lasnik concludes that the "coreferential interpretation" of (3a–b) does not require a linguistic rule relating pronoun and antecedent, but rather arises automatically from the demonstrative use of the pronoun. He does recognize, however, that a rule is needed to account for the contrast between "coreferential" cases like (3a–b) and 'His mother loves John', on the one hand, and noncoreferential cases like (2a–b), on the other. Thus, he postulates a rule of "noncoreference." Evans's criticisms of this view in essay 8 are powerful and persuasive.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

versial assumption that there are no nonrigid singular terms. For if D is a nonrigid antecedent of a rigid occurrence of a pronoun, then the reference of pronoun and antecedent will come apart at certain worlds and ⌜D loves *his* mother⌝ will be assigned the wrong (modal) truth conditions. It also does not suffice to let the referent of a pronoun at an index be the same as that of its nonrigid antecedent. For example, the truth conditions of (5a–b) are given in (6).

(5a) *The* F is such that it is not a necessary truth that *he* is F.
(5b) *The* F has less influence because next year *he* won't be F.
(6a) $\lambda x[-\Box\, x$ is $F]$ the F.
(6b) $\lambda x[x$ has less influence because next year x won't be $F]$ the F.

If ⌜the F⌝ is a nonrigid singular term, however, Evans's treatment of pronouns will fail to provide these truth conditions no matter whether the occurrences of pronouns in (5) are taken to be rigid or nonrigid.

Thus, Evans's treatment of bound pronouns is only as strong as his claim that genuine singular terms are uniformly rigid. Although his analysis of definite descriptions as (binary) quantifiers is attractive, his wholesale rejection of nonrigid singular terms is inconclusive. His main objection to such terms is that they require an otherwise unnecessary theoretical complication—relativization of reference to times and worlds (187–192). At most, this places a burden of proof on theorists who find nonrigid singular terms in natural language. Since the alleged complication is just an instance of the relativization needed for extensions generally, this burden should not be overestimated. Moreover, simplicity aside, it would seem that we could introduce a Fregean description operator *the** or a functor *( )'s height* by explicitly stipulating that it be a singular-term-forming operator with the obvious semantics. This would reinstate the problem posed by (5), since it seems unlikely that such a stipulation should change the interpretation of pronouns, even though some occurrences of pronouns would then have nonrigid antecedents.

The second problem with Evans's analysis of bound pronouns shows up in propositional attitude ascriptions. On that analysis the subordinate clause in

(7a) *John* became F before Mary thought that *he* was F.

expresses an independent proposition—presumably, one that is true iff Mary believed the proposition expressed by ⌜he was F⌝ in (7a). But what proposition is that? Evans repeatedly indicates that it is not the proposition expressed by ⌜John is F⌝ (89–90, 98/9), and it is clear that he would take (7a) and (7b) to be necessarily equivalent

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(where John = Jones), while allowing the truth values of (7c) and (7d) to differ.

(7b) *Jones* became *F* before Mary thought that *he* was *F*.
(7c) Mary thought that John was *F*.
(7d) Mary thought that Jones was *F*.

This strongly suggests that the analysis provides referents but not senses for occurrences of bound pronouns with singular antecedents. Thus, it would seem that the proposition expressed by any sentence (. . . pronoun . . .) containing such an occurrence must be a Russellian singular proposition (that attributes the property λx . . . x . . . to the referent of the pronoun). But Evans does not recognize Russellian propositions as objects of belief. Moreover, the resulting view of pronouns is incorrect in any case. For if '*John* loves *his* mother' expressed a Russellian proposition that predicated of John the property of being one whose mother John loves, then 'Mary believes that *John* loves *his* mother' would express a proposition that said of John that Mary believes him to have this property. Since this proposition may be true when the proposition that Mary believes that John is an own–mother–lover is false, it is not the proposition expressed by the belief ascription. Thus, the appeal to Russellian propositions, apparently needed for (7a–b), cannot be combined with Evans's analysis of bound pronouns.

In my view, the best way to resolve this conflict is to treat (7a) and (7b) as ascribing belief in a Russellian singular proposition, and to analyze bound pronouns as forming complex predicates of the following sort.[3]

(8a) . . . antecedent . . . pronoun . . .
(8b) λx[. . . x . . . x . . .] antecedent.

Since this conflicts with Evans's analysis of bound pronouns, one must presume that he would favor a different solution. It is not clear what it would be, however, since he does not confront the problem directly.

Fortunately, none of this diminishes the significance of his distinction between pronouns, like those in (3) and (4), which are bound

[3] This proposal extends to nonreflexive pronouns the analysis developed in Nathan Salmon, "Reflexivity," *Notre Dame Journal of Formal Logic*, XXVII (1986): 401–429; and utilized in my "Substitutivity," in J. J. Thomson, ed., *On Being and Saying: Essays for Richard Cartwright* (Cambridge: MIT, 1987); and "Direct Reference, Propositional Attitudes and Semantic Content," *Philosophical Topics*, XV (1987): 47–87. This last article is reprinted along with Salmon's "Reflexivity" in Salmon and Soames, eds., *Propositions and Attitudes* (New York: Oxford, 1988).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

by their antecedents, and pronouns, like those in (1), which are not. As Evans points out, if the pronouns in (1) were bound, then (1a–b) would be equivalent to (9a–b)—which they are not.

(9a) *Few students* are such that *they* both came to the party and had a good time.
(9b) There is *a present* such that if Mary buys *it*, then *it* will be expensive.

Evans maintains that instead of functioning as bound variables, the pronouns in (1) are terms whose referents are fixed by descriptions —'the students that came to the party' and 'the present Mary buys'—recoverable from the clauses containing their antecedents. Thus, (1a–b) have the force of (10a–b).

(10a) Few students came to the party, but the students that came to the party (all) had a good time.
(10b) If Mary buys a present, then the present Mary buys will be expensive.

Evans immodestly dubs pronouns of the sort illustrated in (1) as "E-type pronouns" and insists that they are not equivalent to the descriptions that fix their referents. In his system this follows from the fact that they are referring expressions, and are *therefore* rigid, whereas definite descriptions are not. In addition, he cites the fact that the pronouns in (11a–c) do not have the interpretations of small scope descriptions as showing that they are referentially rigid (132/3).

(11a) John owns *a donkey* and *it* likes carrots though it might not have been the case that *it* liked carrots.
(11b) Boston has *a Mayor* and *he* used to be a Democrat.
(11c) *A man* murdered Smith, but John does not believe that *he* murdered Smith.

Nevertheless, it is clear that E-type pronouns are not, in fact, rigid designators. The leading idea is that an occurrence of such a pronoun is supposed to denote the verifiers of the clause containing its antecedent. Since this clause will typically have different verifiers at different circumstances of evaluation, an occurrence of an E-type pronoun associated with it will denote different things in different circumstances. If this were not so, and if 'he' in (11b) rigidly referred to the actual present Mayor of Boston (call him "Kevin"), then (11b) would be true at an arbitrary circumstance of evaluation *E* in which a lifelong Republican was Mayor of Boston at *E*, provided that Kevin used to be a Democrat in *E*. Since this is incorrect, E-type pronouns are typically not rigid designators.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Although this result conflicts with Evans's own characterization, it is fully in accord with the semantic formalism presented at the end of essay 4. There, an E-type pronoun is analyzed as equivalent to a definite description whose scope in a sentence is (roughly) the largest clause containing the pronoun that does not include the clause containing its antecedent. In effect, E-type pronouns are treated not as terms whose reference is rigidly fixed by descriptions, but as terms whose contents are those of large scope descriptions. This means that, contra Evans, either these pronouns are nonrigid referring expressions, or they are not referring expressions at all.[4]

Clearing up Evans's confusion of rigidity with large scope also has the benefit of removing a mystery regarding the kinds of propositions expressed by sentences containing E-type pronouns. Since these propositions exist even when no objects are denoted by the pronouns, they are not singular propositions. On the rigid designator view, they are also not general propositions, since on this view the contents of E-type pronouns are not those of the descriptions which fix their referents. Thus, on the rigid designator view, there is no natural answer to the question 'What kinds of propositions do sentences containing E-type pronouns express?'. Once these pronouns are seen to be descriptions, however, the propositions expressed by sentences containing them can be seen to be unproblematically general.

When E-type pronouns are analyzed as descriptions, it is also not surprising to discover that sometimes they take small scope.[5]

(12a) Although *a student* broke into my office, no one saw *him* and the janitor suspects that *he* had a key.

(12b) *A student* broke into my office, but I doubt that *he* stole anything.

The natural interpretation of (12a) is one in which the janitor suspects that the student who broke into my office had a key, even if he does not suspect any particular person of having a key. Similarly, the doubt in (12b) is naturally interpreted as a general doubt over and above any doubts I might have about specific individuals. To get these interpretations, the E-type pronouns in (12) must be given

[4] E-type pronouns can be antecedents of bound pronouns, as in 'Few candidates voted, but *they* voted for *themselves*'. If, as Evans says, they are referring expressions, they will provide counterexamples to his analysis of bound pronouns (by virtue of being nonrigid).

[5] Similar examples can be found in George Wilson, "Pronouns and Pronominal Descriptions: A New Semantical Category," *Philosophical Studies*, XLV (1984): 1–30. Although examples involving temporal operators are a little harder to come by, the following seems to be a case: 'As always the boys elected *a team captain*, but this year for the first time *he* is an underclassman'.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

small scope, within that of the propositional attitude verbs. Thus, Evans's analysis of these pronouns as always taking large scope is incorrect.

Another aspect of his analysis which requires modification is his treatment of singular E-type pronouns like those in (13).

(13a) Socrates owned *a dog* and *it* bit Socrates.
(13b) If Socrates owned *a dog*, then *it* was registered with the authorities.
(13c) Every farmer who owns *a donkey* beats *it*.

Evans analyzes these syntactically singular pronouns as semantically singular definite descriptions. Several philosophers have noted that this results in the incorrect claims that (13a–b) fail to be true if Socrates owned more than one dog (at the relevant past time) and that (13c) is untrue if one farmer owns several donkeys.[6]

Evans counters by noting the oddness of (14), which he traces to an implication of uniqueness carried by the pronoun.

(14) Socrates has *a finger* and *it* hurts Socrates.

Evans may well be right in taking this example to show that ⌜NP *V*'s *an F* and *it* is *G*⌝ is not analyzed as ⌜NP *V*'s an *F* that is *G*⌝, which of course carries no implication of uniqueness. He is wrong about the truth conditions of (13), however. At a minimum, (13a–b) should be true in a situation in which Socrates owned several dogs, each of which bit him and was registered; and (13c) should be true if each farmer beats every donkey he owns. One gets this result if, following Stephen Neale,[7] one takes 'it' in (13a–b) to be a nonsingular description of all Socrates's dogs, while leaving it open whether he owned only one, or more than one; and similarly for (13c). If this view is correct, then syntactically singular E-type pronouns are not always semantically singular definite descriptions.[8]

[6] See, for example, Wilson, op. cit.; Peter Geach, *Reference and Generality*, pp. 128–131, and "Quine's Syntactical Insights," in his *Logic Matters* (Berkeley: California UP, 1972), pp. 118/9; and Norbert Hornstein, *Logic as Grammar* (Cambridge: MIT, 1984), pp. 80, 85–88.

[7] *Descriptions* (Cambridge: MIT, forthcoming).

[8] On this account, the oddity of (14) might be attributed to the presumption that Socrates has the normal complement of fingers, not all of which hurt. Alternatively, the pronouns in (13a) and (14) may have different analyses from those in (13b) and (13c). As Neale points out, the latter have the force of neutral definite descriptions, where ⌜G(the$_n x$:Fx)⌝ is true iff ⌜All *F*'s are *G*⌝ is true, but it is not specified whether *F* applies to exactly one object (the singular case) or more than one (the plural case). Since this analysis is much less plausible for (13a), (14), and 'Some farmer who owns *a donkey* beats *it*', these pronouns may well have the force of indefinite descriptions, as Wilson has argued. [On this view, (14) carries a cancellable implicature that Socrates has a single finger, but remains true if Socrates has some fingers that hurt and some that do not.]

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

In sum, there are a number of serious problems with Evans's analysis of E-type pronouns. This does not negate, however, the significance of his pioneering work in setting us on the road to understanding these important anaphoric devices.

In another article in the collection, "Reference and Contingency," Evans connects issues in the theory of reference with philosophical questions about the nature of a priori knowledge. In particular, he tries to explain how it is possible to have a priori knowledge of contingent truths like (15a) and (15b) (where the reference of the name 'Julius' is rigidly fixed by the description 'the F').

(15a) If someone uniquely F's, then the actual F is F.
(15b) If someone uniquely F's, then Julius is F.

In brief, Evans's view is that we can have this a priori knowledge because (15a) and (15b) express the proposition (15c), which is clearly a priori.

(15c) If someone uniquely F's, then the F is F.

There is, of course, an obvious objection. Since the proposition expressed by (15c) is necessary, whereas those expressed by (15a) and (15b) are not, the latter are not identical with the former.

In attempting to answer this objection (which he attributes to Kripke), Evans is forced to maintain that, for any possible world w, if w had obtained, then (15a) and (15b) would have been true if and only if (15c) would have been true. For example, let z be the person who is the unique F-er in the actual world, let y be distinct from z, and let w be a world in which y is the unique F-er. According to Evans, if w had obtained, then (15a) and (15b) would have been true. But this is a mistake. It is the mistake of confusing the question 'Is what S says (as used by us here now) true in some alternative circumstance?' with the question, 'Were S uttered in some alternative situation, would it express a truth?'. Although (15a–b) would have expressed truths had they been uttered in w, what they actually express would have been false had w obtained.

One of the things which leads Evans astray is his habit of treating 'had w obtained' and 'had w been actual' as paraphrases. Thus, he glosses, 'Had w obtained (15a) would have been true' as 'Had w been actual (15a) would have been true'. Consequently, in asking whether y would have been the actual F if a world in which y were the F had been actual, he seems to be asking a sensible question which must be answered in the affirmative. This is incorrect, however. 'Actual' is a rigidifier by virtue of being an indexical like 'now' or 'me'. Thus, to

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ask whether another world would have been actual had it obtained is like asking whether another person is me when he speaks, or whether in the future another time will be now, i.e., this time. Moreover, even if the word 'actual' has a nonindexical use in which a world is actual iff it obtains, (15a) will be necessary on such a use, and thus not an example of the contingent a priori.

In the end, this is what Evans's position comes to. On his view, if S expresses something we know a priori, then for all worlds w, had w obtained S would have been true. To use his terminology, all a priori truths are "deeply necessary." Proponents of the contingent a priori are supposed to take solace in the fact that $\ulcorner\Diamond\text{–}S\urcorner$ may be true even if S is deeply necessary, in which case S is said to be "superficially contingent." According to Evans, $\ulcorner\Diamond P\urcorner$ is true iff for some world w, P is "true$_w$." He denies, however, that P is true$_w$ iff (the proposition expressed by) P would have been true had w obtained. Since this is just what proponents of possible world semantics standardly mean by 'true$_w$', it appears that Evans misinterpreted this semantic apparatus.

At best, he can be seen as redefining the modal operators so that they do not express the intended concepts of necessity and possibility. One cannot say what they do express, however, without explaining what 'true$_w$' is supposed to mean. Not only does Evans fail to do this, he maintains that no explanation is needed on the grounds that the term is purely technical and internal to a semantic theory. Fortunately, we need not pause over this. Since Evans's idiosyncratic use of 'true$_w$' is not tied to our intuitive notions of necessity and possibility, his "technical" reinterpretations are irrelevant to explaining genuine instances of the contingent a priori.

This is a pity, since at one point he offers considerations from which we can easily reconstruct the needed explanation. The key move is to relate knowing, in a world w^*, that P, to knowing of w^* that it has the property of satisfying P. Thus, consider the following pairs:

(16a) Tom knows (in w^*) that: $S \leftrightarrow S$
(16b) Tom knows (in w^*) that: $\lambda w[S(w) \leftrightarrow S(w)]w^*$
(17a) Tom knows (in w^*) that: Actually $S \leftrightarrow S$
(17b) Tom knows (in w^*) that: $\lambda w[S(w^*) \leftrightarrow S(w)]w^*$

Let w^* be the actual world and suppose that in each case (a) is true iff (b) is true. The truth of (16a) then guarantees that Tom knows of w^* that it has the property of being an S-satisfying-world-iff-it-is-an-S–satisfying-world. But, whenever a person knows something of this form, he is in a position to know something of the form indicated in

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(17b).[9] Thus, Tom knows of w^* that it has the property of being an S-satisfying–world-iff-w^*-is-an-S-satisfying-world; and (17a) is thereby true. This shows how a priori knowledge of a necessary truth may be sufficient (in the actual world) for a priori knowledge of a contingent truth. Further, nothing in the analysis calls for any distinction between "deep" and "superficial" necessity (contingency), for any redefinition of 'true$_w$', 'actually', or other modal notions, or for the identification of the propositions expressed by the complements of (16a) and (17a).[10]

Although this analysis explains the contingent a priori status of (15a), it does not extend to (15b), which Evans assimilates to (15a). The crucial fact, as I see it, is that since 'Julius' is a name it is not synonymous with any description, even a rigidified one. In addition, I would maintain, with Keith Donnellan,[11] that knowing that Julius is F (if anyone is) requires knowing of the referent of 'Julius' that he is F (if anyone is). But, then, one cannot acquire the belief that Julius is F (if anyone is) merely by using a description to introduce the name (in the absence of epistemic contact with its referent). Thus, Donnellan's position regarding this example is essentially correct. Although one can know a priori that 'Julius is F' expresses a truth if 'Someone uniquely F's' does so as well, this is not sufficient for knowing that, if someone uniquely F's, then Julius does.

In "Does Tense Logic Rest on a Mistake?" Evans takes up related issues involving temporal semantics. He focuses on sentential operators like 'In the past . . .' whose semantics are typically seen as analogous to those of modal operators like '$\Diamond$'. The extensions of these operators are taken to be functions from the intensions of their sentential arguments to truth values. The extension of '$\Diamond$' at w assigns truth to those intensions which are true at one or more worlds accessible from w; the extension of the past tense operator at t

[9] For example, anyone who knows of y that he has the property, $\lambda x[x$ loves $x]$, of loving oneself can conclude that y has the property, $\lambda x[x$ loves $y]$, of loving y. (Variables are directly referential.)

[10] They cannot be identified, even if we restrict our attention to their roles as objects of belief. Surely, it could have been the case in some w that a person believed the proposition expressed by (our use of) the complement of (16a) without believing the proposition expressed by (our use of) the complement of (17a). This could happen either because the person had no belief about the status of S in (the actual world) w^*, or because he believed S to have different truth values in w and w^*. Since it is possible to believe the proposition expressed by (our use of) the complement of (16a) without believing the proposition expressed by (our use of) the complement of (17a), they are distinct.

[11] "The Contingent A Priori and Rigid Designators," in P. A. French, T. E. Uehling, Jr., and H. K. Wettstein, eds., *Contemporary Perspectives in the Philosophy of Language* (Minneapolis: Minnesota UP, 1979), p. 54.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

assigns truth to those intensions which are true at one or more times earlier than t.

This analysis presupposes that (the proposition expressed by) a sentence may have different truth values at different worlds and times. If such variation were not possible, the intension of a sentence would always assign the same truth value to each world and time, and the modal and temporal operators would be otiose. There are, of course, special cases in which (the proposition expressed by) a sentence is explictly indexed to a particular world or time, as in (18–19).

(18a) N is F in w (where 'w' names a particular world).
(18b) N is F on March 31, 1988.
(19a) Actually N is F.
(19b) Presently (now/today) N is F.

In each of these cases, (the proposition expressed by) the sentence is true at an arbitrary world or time if and only if it is true at the explicitly specified world or time, which in the case of (19) is the world or time of the context. Because of this, application of additional modal or temporal operators to these examples is pointless in a way that application of such operators to (20) is not.

(20) N is F.

What the standard analysis presupposes, then, is that (the proposition expressed by) a sentence like (20) (in a particular context) is nonspecific as to world and time, and may, for that reason, receive different evaluations at different worlds and times.

Although Evans accepts this presupposition regarding worlds, he rejects it regarding times. Following Frege, he holds that an utterance of (20) in a context c expresses the proposition that N is F at t_c (the time of c).[12] Since this proposition is indexed for time but not world, its truth value is fixed over time, but not over worlds. If this is

[12] Evans's main argument for this is that, "If a theory of reference permits a subject to deduce merely that [the proposition expressed by] a particular utterance is now correct, but later will be incorrect, it cannot assist the subject in deciding what to say . . ." (349). But why should this be so? Surely the defender of temporally neutral propositions can maintain that, for all times t, one should assert at t only propositions true at t (even if they are false at other times).

A better argument for essentially the same conclusion is provided by Mark Richard in "Temporalism and Eternalism," *Philosophical Studies*, XXXIX (1981): 1–13. Suppose that at t Mary utters 'It is raining' and at $t + n$ states that she continues to believe everything she asserted at t. If what she asserted at t were a temporally neutral proposition p, then her remark at $t + n$ would seem to require her to believe p at $t + n$, and hence to believe it to be raining then. Since, in fact, the truth of her remark at $t + n$ does not require this, we have reason to think that the proposition asserted at t was not temporally neutral, but rather was the proposition that it was raining at t.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

right, then the parallel between modal and temporal semantics breaks down, and another account of temporal operators is needed.

Evans suggests treating them on the model of (21).

(21) For any time t and sentence S, ⌜Past S⌝ is true at t iff there is an earlier time t' such that were S uttered at t', it would express a proposition that is (actually) true.

Evans sees this proposal as recognizing a new form of sentential embedding. Unlike familiar cases in which "the semantic value which a complex sentence $\Sigma(e)$ has in a given context is a function of the semantic value which the expression e has *in that context,*" Evans's proposal "asserts that the semantic value which the sentence ⌜Past (X)⌝ has in a context is a function of the semantic value which X would have in *another* context" (357).

There are two problems here. First, since expressions have many semantic values (extensions, intensions, propositional contents, etc.), there is no such thing as *the* semantic value of an expression. On Evans's proposal, the truth value of ⌜Past S⌝ at t is not a function of the proposition S expresses at t, but it is also not a function of the proposition S expresses at some other context. Rather, it is a function of a different semantic value of S—namely, the function that maps a context c onto the truth value of the proposition expressed by S in c. The effect of Evans's proposal is to take the extensions of temporal operators at times to map these functions onto truth values.

Second, this analysis cannot accommodate temporal indexicals, as is illustrated by its inability to distinguish the following pairs.

(21a) In the past: the present leader of the Republicans was a Democrat.
(21b) In the past: the leader of the Republicans was a Democrat.
(22a) Thirty days ago: no one thought that I would be on the job tomorrow.
(22b) Thirty days ago: no one thought that I would be on the job the following day.

It has long been known that such examples require indexing extensions of expressions both to the time of the context (in order to interpret temporal indexicals) and to another temporal parameter (to interpret familiar tense operators).[13] From this perspective, the recognition that truth values of propositions do not vary over time

[13] The need to double index extensions was noted by Hans Kamp in unpublished work at UCLA in the mid 1960s. Kamp's results were reported in A. N. Prior, "Now," *Nous,* II (1968): 101–119, and published in Kamp, "Formal Properties of 'Now'," *Theoria,* XXXVII (1972): 227–273. The conceptual difference between the roles played by the two time coordinates was generalized to a distinction between

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

requires a semantic framework that differs both from standard temporal systems and from Evans's proposed substitute.

The trick, as Nathan Salmon[14] has shown, is to specify, for each sentence, a two-stage mapping—first, from a context to a temporally neutral propositional structure; and, next, from an arbitrary time to the temporally specific proposition that results from incorporating that time into the initial propositional structure. Salmon calls the second stage of this mapping the schedule of a sentence with respect to a context and takes the extension of a temporal operator at a time to be a function from the schedule of its sentential argument to a truth value.[15] This is what one needs if one wants to make sense of both tense operators and temporal indexicals, while adhering to the traditional view that propositions retain their truth values over time.

Although Evans mostly ignores indexicals in "Does Tense Logic Rest on a Mistake?," he takes them up in "Understanding Demonstratives." He says little about the advances in indexical semantics pioneered by Kaplan and others, but endeavors instead to rebut the view that indexicals pose intractable problems for Fregeans. For our purposes we may take him as accepting Kaplan's semantic machinery, while adding the claim that indexicals have senses that vary from context to context. Like Fregean senses, the senses imagined by Evans are supposed to be "ways of thinking of objects." Unlike the examples given by Frege, Evans's senses are supposed to be rigid,

---

contexts of utterance and circumstances of evaluation by David Kaplan in "On the Logic of Demonstratives," in French, Uehling, and Wettstein, eds., and in his long-circulated manuscript, "Demonstratives," published in J. Almog, J. Perry, and H. K. Wettstein, *Themes From Kaplan* (New York: Oxford, forthcoming). Double indexing of propositions expressed by sentences [needed for example (22)] was introduced by Mark Richard in "Tense, Propositions, and Meanings," *Philosophical Studies*, XLI (1982): 337–351.

[14] *Frege's Puzzle* (Cambridge: MIT, 1986), pp. 24–44; and "Tense and Singular Propositions," in Almog, Perry, and Wettstein.

[15] For example, the schedule of 'I be on the job' with respect to a context c with t_c as time and *S.S.* as agent is a function that maps an arbitrary time t onto the proposition that *S.S.* be on the job at t. The proposition expressed by an utterance of 'I be on the job' in c is the result of applying this schedule to t_c to get the proposition that *S.S.* be on the job at t_c. It is the schedule of the sentence *w.r.t. c* rather than this proposition that determines the truth value of 'Past (I be on the job)' *w.r.t. c* and an arbitrary t. This compound sentence is true *w.r.t. c* and t iff the schedule of 'I be on the job' *w.r.t. c* maps at least one time t^* earlier than t onto an eternally true proposition.

The difference between tense operators and indexicals in (22a–b) is explained as follows: in both cases, the compound sentence is true *w.r.t.* a context c and time t iff at t' 30 days earlier than t no one believed the proposition expressed by the sentential complement of 'thought' *w.r.t. c* and t'. Since the complement sentence in (22a) contains the indexical 'tomorrow', the proposition it expresses *w.r.t. c* and t' is the proposition that *S.S.* be on the job at $t_c + 1$. Since in (22b) the complement contains the tense operator 'the following day', the proposition it expresses *w.r.t. c* and t' is the proposition that *S.S.* be on the job at $t' + 1$.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

nondescriptive modes of presentation which cannot lack referents.[16] Although Evans tries to read such senses back into Frege, the attempt is unpersuasive. The main emphasis is systematic, however, rather than historical. Evans wants to make room for senses in an account of indexicals which is Fregean in spirit.

In a way, nothing could be easier. As long as one says only that senses of indexicals in particular contexts are primitive and nondescriptive (without saying what they are), the standard arguments against a Fregean account will not apply. It is not clear, however, that there is anything to be gained from such an approach. Evans's senses are not needed to determine reference, since the referent of an indexical in a context is already determined by a linguistic rule—its meaning or character in Kaplan's sense. Nor are senses needed to explain how utterances of sentences which differ only in coreferential indexicals can differ in informativeness for a fully competent speaker.[17] The main hope seems to be that introducing senses will block problematic results involving propositional attitudes, most notably substitutivity in attitude ascriptions. But even this is doubtful.

An appeal to senses will block substitutivity only in cases in which the indexicals undergoing substitution have different senses. Thus, in order to have any consequences about when substitutivity is permitted and when it is not, a theory must indicate for which indexicals i, j and contexts c, c' the sense of i in c is the same as the sense of j in c'. Since Evans mostly ignores this issue, few results about substitutivity are forthcoming.

[16] Evans's argument for such senses is given on pages 294–296, and can be reconstructed as follows: since the sense of a term is a way of presenting a referent, a theory of meaning which includes the term will give its sense by specifying its referent (in a manner which displays or uses the sense of the term without stating that it is the sense). In the case of a syntactically simple singular term, this job will be done by an axiom of the form 'The referent of 't' = t'. Since no such axiom can be true unless t has a referent, every such singular term that has a sense must have a referent. Thus, there cannot be meaningful, syntactically simple singular terms that lack referents.

In addition, Evans holds that "the sense of a singular term is a way of thinking about a particular object: something that obviously could not exist if that object did not exist to be thought about" (295). Here, he seems to hold that a nondescriptive sense that presents an object o could not have existed if o had not. This claim, which is given no independent justification, goes well beyond the argument's conclusion. Moreover, the conclusion itself is easily avoided. Even a Davidsonian can state the axiom for a term in a way that does not presuppose a referent—e.g., 'For all x ('t' refers to x iff $x = t$)'. But then nothing requires terms with senses to have referents.

In addition to being poorly motivated, Evans's nondescriptive senses are rather mysterious. Although they are supposed to determine reference, he does not say how they do this. Nor is it clear what grasping a sense will consist in, since agents in identical psychological states must be able to grasp different senses.

[17] See Kaplan, "Demonstratives," especially section XVII; and my "Substitutivity," especially section II.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

There are exceptions, however. Evans agrees with Frege that 'Yesterday was F' expresses the same thought on $d + 1$ as 'Today is F' does on d, from which he concludes that the sense of 'yesterday' on $d + 1$ is the same as the sense of 'today' on d. But this is tantamount to recognizing that the reference of these indexicals determines their sense, from which it follows that their senses are incapable of doing semantic work not done by the referents alone. In particular, they have no effect on substitutivity or the ascription of propositional attitudes to individuals. For example, suppose Pierre sincerely and reflectively utters 'Today is F' at 11:58 PM and 'Yesterday was not F' three minutes later. Evans's brand of Fregeanism is committed to characterizing this case in exactly the same way that direct reference theorists do—namely, as being one in which Pierre unknowingly asserts and believes contradictory propositions without being guilty of any logical or linguistic error. If other indexicals are treated on the model of 'yesterday' and 'today', then the result generalizes.[18]

In principle, there seems to be little reason not to extend this treatment to 'I', 'you', and 'he'. If utterances of 'Today is F' and 'Yesterday was F' on different days can express the same proposition, then utterances of 'I am F', 'You are F', and 'He is F' by different people should be able to do so as well. This is, however, incompatible with Frege's view, defended by Evans, that each person associates a sense with 'I' which is not graspable by anyone else.[19] This appeal to logically private senses creates nothing but problems. For example, given standard assumptions about the attitudes, one gets the results that only I can believe or assert that I have the property F, that no one else can even report my belief or assertion in indirect discourse, and that I cannot report others' attitudes about me by saying "They assert (believe) that I am F." These results conflict with our most fundamental practices of expressing and reporting propositional attitudes.

[18] Evans appears to follow Frege in treating 'here' and 'there' on the model of 'yesterday' and 'today' (310/1). Frege suggests a further generalization when he says, "The pointing of fingers, hand movements, glances, may belong here too"; in "The Thought: A Logical Inquiry," in P. F. Strawson, ed., *Philosophical Logic* (New York: Oxford, 1967), p. 24.

[19] It is not obvious why a Fregean should defend such a view. Because sense is supposed to determine reference, the Fregean needs to prevent the same indexical ('I', 'today', 'here') from being associated with the same sense in different contexts (with different agents, times, places). With 'I' this is accomplished by making the sense a person attaches to it inaccessible to others (no matter what their psychological state). With 'today' (or 'here') it is accomplished by making its sense at a time (place) accessible at other times (places), but only as the sense of a different expression. Since the semantic problem is the same in the two cases, it is not clear why the cases should be treated differently.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

They also illustrate a dilemma that Evans refuses to face. If one generalizes Frege's treatment of 'today' and 'yesterday' by identifying the senses of indexicals in contexts in which their referents are the same, then one eliminates the rationale for senses and ends up with a notational variant of direct reference theories. If, on the other hand, one generalizes Frege's treatment of 'I' by positing different senses for coreferential indexicals, then one comes into conflict with our basic practices of using indexicals in the ascription of attitudes.

In sum, there is much in *Collected Papers* to argue with. There is, however, also much to be learned from coming to grips with Evans's challenging and often illuminating discussions.

SCOTT SOAMES

Princeton University

Collected Papers. GARETH EVANS. New York: Oxford University Press, 1985. 407 p. $39.95.*

The last paper in this collection, "Molyneux's Question," which is a first draft that Evans would not have published as it stood, is path-breaking interdisciplinary work. It is not an essay in applied philosophy, in the philosophy *of* psychology; nor is it a plea for the replacement of philosophy by psychology, or a philosophical discussion to which psychology is in a general way relevant. The question is whether a man born blind, who has learned to distinguish cubes from spheres and who now recovers his sight, would be able to tell on the strength of sight alone which were which. What becomes clear in Evans's richly referenced development of the issues is that even to get them properly stated requires one to draw upon both disciplines. This is both because the shape concepts that Molyneux's question uses have to be explicated in terms made available by the psychological study of perception, and because the background issue of the epistemological relations between the various senses cannot be seriously addressed without appeal to psychological work.

I shall briefly mention three of the widely known papers in the collection before returning to this essay, which is published here for the first time. At the end of this review, I shall indicate its relation to Evans's "Things Without the Mind," also included in this book.

* An earlier version was presented at a conference on Spatial Representation at King's College, Cambridge, and I am indebted to participants for their comments. Thanks to Roz McCarthy, to whom I owe the neuropsychological references.

0022-362X/89/8603/155–163 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

I.

In "Identity and Predication," Evans begins by elaborating a constraint upon interpretation: that we can discern predication of *A*s in sentences of a given kind, only if grasp of them requires mastery of the identity conditions of *A*s. Evans has in mind here the way in which the assent conditions of sentences involving complex predicates can require sensitivity to the boundaries of an object. For example, sensitivity to the assent conditions of 'there is an *A* here which is *F* and an *A* here which is not *F*' requires sensitivity to the boundaries of *A*s; and sensitivity to the assent conditions of 'there is an *A* here which was *F* and which will be *G*' requires sensitivity to the boundaries of the object over time. This point is then brought to bear on W. V. O. Quine's various nonstandard interpretations of 'gavagai', which are held to cut either too finely or too coarsely. Those which cut too coarsely, such as 'rabbiteth', cannot interpret complex predications of the type just indicated; those which cut too finely leave us with a situation in which, for instance, "different parts of the same rabbit are indistinguishable by the predicates of the language" (43). In this situation, we are discussing predicative structure whose discernment is not licensed by Evans's constraint.

"Semantic Structure and Logical Form" aims to characterize a notion of "structurally valid inference," inference depending only upon structure, as opposed to inference whose validity is dependent upon the semantic assignments made to particular primitive terms, including the logical constants. Thus, a "structurally valid inference" would be the transition from 'John is a tall man' to 'John is a man'. To explain the distinction, Evans appeals to the semantic categories of a model theory for the language; categories being established by their combining illumination with generality in explaining patterns of inference. In an appendix, Evans tries to put in perspective this contrast between types of validity and suggests a way of identifying the logical constants.

"Semantic Theory and Tacit Knowledge" proposes "that we construe the claim that someone tacitly knows a theory of meaning as ascribing to that person a set of dispositions—one corresponding to each of the expressions for which the theory provides a distinct axiom" (328/9). This totality of dispositions is simultaneously characterized by the axioms, and each such disposition is taken to have its neural ground. Evans stresses his view of semantic knowledge as an inferentially insulated input system, rather than as part of a central system in which information is freely accessed.

These papers each give canonical statements of views that are of central interest, and I do not propose to comment on them here, beyond remarking that "Identity and Predication" seems particu-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

larly likely to be of permanent interest. Instead, I turn to look in more detail at "Molyneux's Question."

II.

On one view, we have two concepts of 'sphere', 'straight line', 'between', and so on, tactile and visual respectively. Someone who advocates this view might accept that both sight and touch give information about the very same network of spatial properties: there is only one system of spatial relations. But he is, in effect, using a Fregean test for individuating concepts, ways of thinking of spatial properties. He takes it that, if the tactile way of thinking of cubicity is the very same as the visual way, then there must be no room for doubt as to whether it is the same property one is seeing as can be apprehended by touch. We do not ordinarily have any misgivings here; but the point of focusing on Molyneux's newly sighted man, recovering from blindness, is to bring out that *that* may be merely the upshot of the pervasive correlations between sight and touch to which we are accustomed. The newly sighted man, innocent of any experience of those correlations, could see a block *as* cubic, but, even if he trusts his sight, could wonder whether it has that property of cubicity which he apprehends through touch. This issue is independent of problems about the degeneration of the visual system when it goes unused. It is also independent of the question of nativism. If we have different concepts of tactile and visual cubicity, these might be learned or innate, and the correlations between them might be learned or innate. So, if a newly sighted man actually did confidently use his old vocabulary for shapes in classifying seen objects, this might be the upshot merely of an innate correlation between a pair of distinct though innately given concepts; it would not establish positively that there is here a single conceptual capacity.

The aim of Evans's paper is to establish the unitary character of our grasp of a shape concept. He begins by introducing the notion of *egocentric* space. In effect, perception is taken to locate objects in a coordinate frame whose axes are given by 'up', 'down', 'right', 'left', 'in front', and 'behind', and whose origin is designated by 'here'. The strategy is to show that this frame is used by both sight and touch, and that our ordinary shape concepts are held to be partly constituted as responses to that information. But, since it is the same spatial information supplied by the two senses, our ordinary shape concepts are then evidently univocal: it is the same conceptual capacity that is being exercised when one applies 'cubic' on the basis of sight as when one applies it on the basis of touch.

To show this coordinate frame is common to the spatial senses, Evans stresses the connection between egocentric space and what is sometimes called behavioral space: "having the perceptual informa-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tion [as to the location of an object] at least partly consists in being disposed to do various things" (383). To hear a sound as coming from the right *is* to be immediately capable of pointing toward it, and of moving closer to it or further away from it. And a parallel point seems to hold for tactual/kinaesthetic perception, as Evans points out: "when he uses his hand, the blind man gains information whose content is partly determined by the disposition he has thereby exercised—for example, that, if he moves his hand forward such-and-such a distance and to the right, he will encounter the top part of a chair. . . . [When he has synthesized a sequence of haptic perceptions into a unitary representation] [e]ach bit of the information is directly manifestable in his behaviour" (389). The same is true, Evans holds, for sight. The crucial point is that, as he puts it, "There is only one behavioural space." All the spatial senses feed into a single repertoire of movements.

III.

As we shall see, the connection between shape concepts and egocentric space can be called into question; but first, let us put this approach into context. When Evans writes: "Auditory input, or rather, the complex property of auditory input which codes the direction of sound, acquires spatial content for an organism by being linked with behavioural output in an advantageous way" (385), he places his account within a developing tradition of "naturalizing" approaches to representational content. The view is that, in order to find what a particular state or process of an organism represents, one has to look at what it is for, at the reasons why there have evolved creatures capable of going into those states as responses to external stimuli. What makes it so that perception has spatial content is, on this account, just that the reason why those perceptual systems have been selected is their value in the production of spatial behavior, in guiding the direction and distance of movement. There is no mystery, on this approach, about how representational content can be part of the natural world.

What we have so far is an assurance that hearing and touch acquire spatial content in the same kind of way, because of their role in directing behavior. It also seems plausible that vision may acquire its spatial content in just that way. All the spatial senses, then, acquire spatial content through their role as input into the direction of a single behavioral repertoire. If this is so, then Evans will claim to have secured his basis for the claim that a unitary conceptual ability is being exercised when we apply a shape concept indifferently on the strength of touch and vision.

This strategy depends, however, on the assumption that, if sight and touch both deliver the same kind of egocentric spatial informa-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tion as to the whereabouts of objects, then it will be apparent to the subject that it is the same information that he is receiving. The assumption is that this information will meet a broadly Fregean constraint: that it will be transparent to the subject that different senses are providing him with the very same information about the location of a thing, if they are doing so. The problem is that we have no guarantee that the kind of naturalized content I indicated will meet this Fregean constraint. On the face of it, it could be that vision and touch direct behavior in ways that are in all the relevant respects similar, so that they can be viewed as both supplying the very same sort of egocentric content; and yet it might be that this was not something apparent to the percipient, that it was, on the contrary, something that could be discovered only by experiment.

As I said, Evans writes as if the crucial point is that "[t]here is only one behavioural space" (390). Here he appeals to work showing that compensation for visual rearrangement—induced by prisms shifting the apparent locations of objects, for instance—produces changes in auditory localization. That this is true, however, does not mean that it is apparent to the percipient that it is true—indeed, if it did, the experimental work would not be so informative. And if, despite their similar roles in the direction of behavior, visual and tactual spatial information are *not* found manifestly identical by the percipient, then spatial concepts constituted as a response to that information will not be univocal, by Frege's test. It will be possible for the percipient to doubt whether the property of cubicity which he accesses through sight is the same as the property he accesses through touch.

There is, indeed, some immediate plausibility in the view that the sameness of perceptual space across modalities is transparent to the subject. But once we see that this cannot be explained by the sameness of naturalized content across modalities, other possibilities are opened. Most radically, it might be held that here we must appeal to the role of language in serving as a bridge across the modalities. It might be that we have to appeal to the subject's own grasp of egocentric spatial concepts as the cross-modal bridge. We should then have a "top-down" rather than a "bottom-up" approach: at the level at which we have transparency, the perceptual contents will be individuated in part by the conceptual thinking to which they are input, and by how they are input to it. We should simply have lost sight of the naturalized content favored by Evans, and his proposal to explain transparent unity in terms of it and its connections with behavior.

IV.

Suppose we set aside this difficulty of principle. Is it correct in any case to view simple shape concepts as responses to egocentric information, whether or not it is transparently delivered indifferently by

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

sight and touch? Egocentric space is essentially a space within which the positions of objects are identified: its whole point is to say just where this or that thing is with respect to the percipient. And it is often said that we need a distinction between the ways in which we find *where* things are, and the ways in which we find *what* they are. The capacities can break down independently of each other.

An illustration is provided by patients suffering from "visual disorientation," classically defined as the inability to localize the positions and distances of objects by sight alone.[1] In effect, these patients receive visual spatial information only in an object-centered coordinate frame; they are not given locations in egocentric space. They retain their capacity to recognize the shapes of objects on the strength of sight, and indeed they can recognize simple geometrical shapes displayed on a screen.[2] But they do not receive egocentric information. Thus, they have no sense of the relative distances of objects: "Some of the men walked repeatedly up against walls, as they had no idea how near they were to them, though they saw them distinctly."[3] They cannot tell the relative sizes of objects; thus, one patient could discriminate spoons by sight, but had to resort to touch to find which were the teaspoons.[4] The relative locations of objects cannot be established by sight: which objects are right, left, up, or down from each other cannot be seen, though objects themselves can. (So, although individual words can be recognized, reading presents serious difficulties.[5]) Thus, vision here gives only object-centered spatial information: the egocentric frame has been lost. It has been established that this is not a matter of any motor defect or reaching disability.[6] The opposite phenomenon also occurs: the "blindsighted" patient studied by L. Weiskrantz[7] could identify the egocentric locations of objects without having any knowledge of their shapes. These data point toward an independence of the mechanisms for shape recognition and egocentric location.

Since shape recognition and the capacity for egocentric location can come apart like this, it is not clear why the existence of a single egocentric space, common to sight and touch, should imply that the capacity to recognize shapes is also cross-modal. Evans does note the

[1] Gordon Holmes, "Disturbances of Visual Space Perception," *British Medical Journal* (1919): 230–233.
[2] R. B. Godwin-Austen, "A Case of Visual Disorientation," *Journal of Neurology, Neurosurgery and Psychiatry*, XXVIII (1965): 453–458.
[3] Holmes, *op. cit.*, p. 231.
[4] Godwin-Austen, *op. cit.*, p. 455.
[5] *Ibid.*
[6] Marvin Cole, Henry S. Schutta, and Elizabeth K. Warrington, "Visual Disorientation in Homonymous Half-Fields," *Neurology*, XII (1962): 257–263.
[7] *Blindsight* (New York: Oxford, 1986).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

possibility of a creature that can discriminate shapes, but not locate and orient toward objects in visual space. His comment on this is an insistence that there is no such thing as experience—conscious experience—of shapes otherwise than *as* oriented in egocentric space (395/6). It is not at all obvious that this is the right way to describe the above cases. But suppose it is. Then experience of shapes must be the joint upshot of, among other things, two distinct capacities: the capacity to locate objects in egocentric space, to tell *where* they are, and secondly, the capacity to tell *what* shapes they are. We do not have experience of shapes unless both of these capacities are operative. It by no means follows that, if one of these capacities is cross-modal, the other must be cross-modal too.

Certainly there are some shapes, recognition of which need not be cross-modal, on any account. For example, the ability to identify faces, or a particular type of tree, by shape need not be transferable from vision to touch, even if there is only a single egocentric space. Why should it be any different for simpler shapes, such as the cube or the sphere? Even if there is a single egocentric space, the capacity to recognize cubicity might remain modality specific.

It is tempting to think that the notion of a *configuration* of objects provides, as it were, a bridge between apprehension of egocentric location and apprehension of shape. But the capacity to find the the several egocentric locations of a number of objects is one thing, and the capacity to perceive them as *configured* in a certain way is another. We have so far, it seems to me, simply no reason to suppose that the ability to recognize a particular configuration must be cross-modal, even if there is only a single egocentric space.

V.

The last problem I want to raise for Evans's approach to Molyneux's question is this. Suppose we agree that shape concepts are partially constituted as responses to perceptual information, that the relevant level of perceptual information is constant across the senses, and that this is transparent to the subject. The problem is whether this sameness, at the level of how things cross-modally seem, could be enough to give the subject knowledge that it is the same property that he is accessing through sight as through touch. For there to be a unitary conceptual capacity here, the subject must know that it is the same property he has knowledge of through sight as through touch. Otherwise, he could doubt whether it was the same property, and we should have different concepts.

The underlying problem here is that grasp of a shape concept cannot be viewed simply as a matter of one's capacity to respond appropriately to perceptual input; there is also a background sense of "what one is about" *organizing* one's use of the concept. Conse-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

quently, the question whether a unitary conceptual capacity is being exercised in connection with both sight and touch cannot be resolved merely by looking at the character of the perceptual input to which one's use of the concept is a response. We have also to look at the character of this background "organizing" conception to find whether that really does permit the application of the same concept indifferently on the basis of sight and touch.

The question how to characterize this background "organizing" conception is inordinately difficult; but one way to attack it is shown in "Things Without the Mind," an extended discussion of chapter 2 of Strawson's *Individuals*. This landmark essay does not deal directly with the relations between sight and touch. Rather, he is explicitly concerned with Strawson's question, whether there could be a creature that thought in terms of objective particulars, but not in terms of spatial things; and, like Strawson, he orients the question in terms of the conceptual repertoire possessable by a creature relying solely on hearing (it being assumed that hearing does not of itself provide for spatial awareness). Evans's discussion draws out many aspects of the way in which our most elementary spatial thinking is organized, and might be taken as exemplary for an investigation of the way we ordinarily think about the relations between sight and touch. But we might also view this discussion, and Strawson's, as illuminating an issue that they did not mean to emphasize: namely, as articulating in depth the ways in which the kind of consciousness of the world provided by ordinary binaural hearing is conceptually and epistemically dependent upon sight and touch. The problem which remains, one aspect of which we have been discussing, is to articulate similarly the dependencies between ordinary sight and touch.

These essays illuminate every topic on which they touch. I have focused on "Molyneux's Question" and "Things Without the Mind," because of the way in which they open up the topic of the epistemic and conceptual relations between different sensory modalities. For anyone interested in a new direction of approach to epistemology and the theory of content, there could not be a better starting point than these papers, with their distinctive weave of philosophy and psychology.

JOHN CAMPBELL

New College/Oxford

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

NEW BOOKS: ANTHOLOGIES

BARLOW, MICHAEL, and CHARLES A. GERGUSON, eds.: *Agreement in Natural Language: Approaches, Theories, Descriptions.* Chicago: University Press, 1988. xiv, 355 p. Cloth $35.00.

BAYEFSKY, ANNE, ed.: *Legal Theory Meets Legal Practice.* Edmonton, Canada: Academic, 1988. xxvi, 369 p.

BEHLER, ERNST, and JOCHEN HÖRISCH, eds.: *Die Aktualität der Frühromantik.* Paderborn: Schöningh, 1988. 264 p.

BENJAMIN, ANDREW, ed.: *Post-Structuralist Classics.* New York: Routledge, 1988. viii, 273 p. Cloth $75.00.

BROMLEY, DAVID G., ed.: *Falling from the Faith: Causes and Consequences of Religious Apostasy.* Newbury Park, CA: Sage, 1988. 266 p. Cloth $35.00, paper $16.95.

BROWN, HARVEY R., and ROM HARRE, eds.: *Philosophical Foundations of Quantum Field Theory.* New York: Oxford, 1988. 190 p. Cloth $49.95.

BURKHARDT, FREDERICK, and FREDSON BOWERS, eds.: *The Works of William James: Manuscript Lectures.* Cambrdige: Harvard, 1988. lxiii, 686 p. Cloth $70.00. *The Works of Williams James: Manuscript Essays and Notes.* Cambridge: Harvard, 1988. xlviii, 550 p. Cloth $50.00.

CHISHOLM, RODERICK M., and JOHANN C. MAREK, eds.: *Franz Brentano: Über Ernst Machs* Erkenntnis und Irrtum. Amsterdam: Rodopi, 1988. 244 p.

CLARK, PETER, and CRISPIN WRIGHT, eds.: *Mind, Psychoanalysis, and Science.* New York: Blackwell, 1988. xiv, 370 p. Cloth $49.95.

CONNOLLY, JOHN M., and THOMAS KEUTNER, eds.: *Hermeneutics vs. Science? Three German Views: Wolfgang Stegmüller, Hans-Georg Gadamer, Ernst Konrad Specht.* Notre Dame: University Press, 1988. vii, 128 p. Cloth $15.95, paper $7.95.

DAGGY, ROBERT E., et al.: *The Merton Annual, vol. 1.* New York: AMS, 1988. x, 385 p. Cloth $42.50.

DANCY, JONATHAN, et al., eds.: *Human Agency: Language, Duty, and Value.* Stanford: University Press, 1988. xi, 308 p. Cloth $35.00.

DIAMOND, IRENE, and LEE QUINBY, eds.: *Feminism and Foucault: Reflections on Resistance.* Boston: Northeastern, 1988. xx, 246 p. Cloth $37.50, paper $14.95.

DRONKE, PETER, ed.: *A History of Twelfth-century Western Philosophy.* New York: Cambridge, 1988. xi, 495 p. Cloth $69.50.

ELSTER, JON, and RUTH SLAGSTAD, eds.: *Constitutionalism and Democracy.* New York: Cambridge, 1988. viii, 359 p. Cloth $39.50.

GAMBETTA, DIEGO, ed.: *Trust: Making and Breaking Cooperative Relationships.* New York: Blackwell, 1988. xii, 246 p. Cloth $34.95.

GAVIN, WILLIAM J., ed.: *Context over Foundation: Dewey and Marx.* Boston: Reidel, 1988. 259 p. Cloth $69.00.

GEISMANN, GEORG, and KARLFRIEDRICH HERB, eds.: *Hobbes über die Freiheit.* Wurzburg: Königshausen, 1988. 228 p.

GEISMANN, GEORG, and HARIOLF OBERER, eds.: *Julius Ebbinghaus: Philosophie der Freiheit: Praktische Philosophie: 1955–1972.* Bonn: Bouvier, 1988. xii, 467 p.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

GERBER, WILLIAM, ed.: *A Philosopher's Harvest: The Philosophical Papers of Isaac Frank.* Washington, Georgetown UP, 1988. xii, 244 p. Cloth $24.95, paper $12.95.

GERGEN, MARY MCCANNEY, ed.: *Feminist Thought and the Structure of Knowledge.* New York: University Press, 1988. xix, 200 p. Cloth $35.00.

GILLESPIE, MICHAEL ALLEN, and TRACY B. STRONG, eds.: *Nietzsche's New Seas: Explorations in Philosophy, Aesthetics, and Politics.* Chicago: University Press, 1988. v, 240 p. Cloth $24.95.

GOODMAN, MICHAEL F., ed.: *What is a Person?* Clifton, NJ: Humana, 1988. ix, 325 p. Cloth $29.50.

GRIFFIN, DAVID RAY, ed.: *The Reenchantment of Science.* Albany: SUNY, 1988. xv, 173 p. Cloth $34.50, paper $10.95.

GRISWOLD, CHARLES, L., ed.: *Platonic Writings/Platonic Readings.* New York: Routledge, 1988. 321 p. Cloth $47.50, paper $13.95.

GRÜNBAUM, ADOLF, and WESLEY C. SALMON, eds.: *The Limitations of Deductivism.* Berkeley: California UP, 1988. xxv, 168 p. Cloth $32.50.

HELMAN, DAVID H., ed.: *Analogical Reasoning: Perspectives of Artificial Intelligence, Cognitive Science, and Philosophy.* Boston: Kluwer, 1988. vii, 428 p. Cloth $99.00.

HERTZBERG, LARS, and JUHANI PIETARINEN, eds.: *Philosophy of History and Culture.* Leiden: E. J. Brill, 1988. 176 p.

HOFFMAN, R. JOSEPH, and GERALD A. LARUE, eds.: *Biblical versus Secular Ethics: The Conflict.* Buffalo: Prometheus, 1988. 191 p. Cloth $22.95.

HURAY, PETER LE, and JAMES DAY, eds.: *Music and Aesthetics in the Eighteenth and Early-nineteenth Centuries.* New York: Cambridge, 1988. xiv, 399 p. Paper $24.95.

KIECKHEFER, RICHARD, and GEORGE D. BOND, eds.: *Sainthood: Its Manifestations in World Religions.* Berkeley: California UP, 1988. xii, 263 p. Cloth $38.50.

KING, NANCY M. P., LARRY R. CHURCHILL, and ALAN W. CROSS, eds.: *The Physician as Captain of the Ship.* Boston: Reidel, 1988. xv, 254 p. Cloth $59.00.

KNUUTTILA, S., ed.: *Modern Modalities: Studies of the History of Modal Theories from Medieval Nominalism to Logical Positivism.* Boston: Kluwer, 1988. xiv, 345 p. Cloth $99.00.

KOLENDA, KONSTANTIN, ed.: *Organizations and Ethical Individualism.* New York: Praeger, 1988. xxi, 172 p. Cloth $37.95.

KRETZMANN, NORMAN, ed.: *Meaning and Inference in Medieval Philosophy: Studies in Memory of Jan Pinborg.* Boston: Kluwer, 1988. xii, 400 p. Cloth $129.00.

KRETZMANN, NORMAN, ANTHONY KENNY, and JAN PINBORG, eds.: *The Cambridge History of Later Medieval Philosophy.* New York: Cambridge, 1988. xiv, 1035 p. Paper $27.95.

LEITES, EDMUND, ed.: *Conscience and Casuistry in Early Modern Europe.* New York: Cambridge, 1988. ix, 269 p. Cloth $39.50.

LIST, ELIZABETH, and ILJA SRUBAR, eds.: *Alfred Schütz: Neue Beiträge zur Rezeption seines Werkes.* Amsterdam: Rodopi, 1988. 348 p.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

McLAUGHLIN, BRIAN P., and AMELIA OSKENBERG RORTY, eds.: *Perspectives on Self-Deception.* Berkeley: California UP, 1988. vi, 558.

McMURRIN, STERLING M., ed.: *The Tanner Lectures on Human Values: VIII.* Salt Lake City: Utah UP, 1988. xii, 333 p. Cloth $25.00.

MENDUS, SUSAN, ed.: *Justifying Toleration: Conceptual and Historical Perspectives.* New York: Cambridge, 1988. vii, 260 p. Cloth $39.50.

MORRIS, THOMAS V., ed.: *Divine and Human Action: Essays in the Metaphysics of Theism.* Ithaca: Cornell, 1988. vi, 366 p. Cloth $42.50, paper $13.95.

MORRIS, THOMAS V., ed.: *Philosophy and the Christian Faith.* Notre Dame: University Press, 1988. xiv, 300 p. Cloth $31.95.

MURPHY, MURRAY G., and IVAR BERG, eds.: *Values and Value Theory in the 20th Century America: Essays in Honor of Elizabeth Power.* Philadelphia: Temple, 1988. vi, 301 p. Cloth $34.95.

NEWMARK, KEVIN, ed.: *Phantom Proxies: Symbolism and the Rhetoric of History.* New Haven: Yale, 1988. vii, 288 p. Paper $13.95.

NOVAK, JOSEF, ed.: *On Masaryk: Texts in English and German.* Amsterdam: Rodopi, 1988. 398 p.

ODERGARD, DOUGLAS, ed.: *Ethics and Justification.* Edmonton, Canada: Academic, 1988. 298 p.

OTTO, HERBERT R. and JAMES A. TUEDIO, eds.: *Perspectives on Mind.* Boston: Reidel, 1988. ix, 420 p. $79.00.

PETTIT, PHILIP, et al., eds.: *Metaphysics and Morality.* New York: Blackwell, 1988. viii, 212 p. Cloth $45.00.

PITT, JOSEPH C., ed.: *Theories of Explanation.* New York: Oxford, 1988. 222 p. Cloth $32.00, paper $11.95.

PORTER, ROY, and MIKULAS TEICH, eds.: *Romanticism in National Context.* New York: Cambridge, 1988. x, 353 p. Cloth $54.50, paper $17.95.

REESE, WILLIAM L., ed.: *The Best in the Literature of Philosophy and World Religions.* New York: Bowker, 1988. xxx, 801 p. Cloth $75.00.

ROSENBAUM, ALAN S., ed.: *Constitutionalism: The Philosophical Dimension.* New York: Greenwood, 1988. xii, 263 p. Cloth $45.00.

ROSENBERG, ALAN, and GERALD E. MEYERS, eds.: *Echoes from the Holocaust: Philosophical Reflections on a Dark Time.* Philadelphia: Temple, 1988. xvi, 453 p. Cloth $34.95.

RUSE, MICHAEL, ed.: *But Is It Science?* Buffalo: Prometheus, 1988. 406 p. Cloth $23.95.

SCHNEIDER, JEAN, and MONIQUE LEGER-ORINE, eds.: *Frontiers and Space Conquest.* Boston: Kluwer, 1988. viii, 255 p. Cloth $43.00.

SCHURZ, GERHARD, ed.: *Erklaren und Verstehen in der Wissenschaft.* Munich: Oldenbourg, 1988. 341 p.

SEANOR, DOUGLAS, and N. FOTION, eds.: *Hare and Critics: Essays on Moral Thinking.* New York: Oxford, 1988. viii, 307 p. Cloth $49.95.

SEBBA, GREGOR: *The Dream of Descartes.* RICHARD A. WATSON, ed. Carbondale: Southern Illinois UP, 1987. (*Journal of the History of Philosophy* Monograph Series.) xiv, 75 p. Paper $12.95.

SHANKER, S. G., ed.: *Gödel's Theorem in Focus.* New York: Routledge, 1988. ix, 261 p. Cloth $59.95.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

SILVERMAN, HUGH J., et al., eds.: *The Horizons of Continental Philosophy: Essays on Husserl, Heidegger, and Merleau-Ponty.* Boston: Kluwer, 1988. xvii, 304 p. Cloth $78.00.

TYMIENIECKA, ANNA-TERESA, ed.: *Analecta Husserliana: The Yearbook of Phenomenological Research, vol. XXII: Morality within the Life- and Social World.* Boston: Reidel, 1987. xv, 586 p. $98.00.

TYSON, RUEL W. JR., JAMES L. PEACOCK, DANIEL W. PATTERSON, eds.: *Diversities of Gifts: Field Studies in Southern Religion.* Champaign: Illinois UP, 1988. xiv, 218 p. Cloth $21.95.

UGRINSKY, ALEXEJ, eds.: *Friedrich von Schiller and the Drama of Human Existence.* New York: Greenwood, 1988. viii, 205 p. Cloth $39.95.

WOOD, DAVID, and ROBERT BERNASCONI, eds.: *Derrida and Differance.* Evanston: Northwestern, 1988. xiii, 98 p. Cloth $24.95, paper $11.95.

ZANER, RICHARD M., ed.: *Death: Beyond Whole Brain Criteria.* Boston: Kluwer, 1988. ix, 276 p. Cloth $69.00.

NEW BOOKS: TRANSLATIONS

AL-FAYYUMI, SAADIAH BEN JOSEPH: *The Book of Theodicy: Translation and Commentary on the Book of Job.* L. E. Goodman, trans. New Haven: Yale, 1988. xvii, 481 p. Cloth $75.00.

BERGSON, HENRI: *Matter and Memory.* N. M. Paul and W. D. Palmer, trans. New York: Zone, 1988. 284 p. Cloth $21.95.

CERTEAU, MICHEL DE: *The Writing of History.* Tom Conley, trans. New York: Columbia, 1988. xxviii, 368 p. Cloth $40.00.

DELEUZE, GILLES: *Bergsonism.* Hugh Tomlinson and Barbara Habberjam, trans. New York: Zone, 1988. 131 p. Cloth $19.95.

DE MOLINA, LUIS: *On Divine Knowledge: Part IV of the Concordia.* Alfred J. Freddoso, trans. Ithaca: Cornell, 1988. xii, 286 p. Cloth $34.95.

ECO, UMBERTO: *The Aesthetics of Thomas Aquinas.* Hugh Bredin, trans. Cambridge: Harvard, 1988. xi. 287 p. Cloth $30.00, paper $14.95.

HEIDEGGER, MARTIN: *Hegel's Phenomenology of Spirit.* Parvis Emad and Kenneth Maly, trans. Bloomington: Indiana UP, 1988. xviii, 156 p. Cloth $25.00.

HODGSON, PETER C., ed.: *Hegel: Lectures on the Philosophy of Religion: The Lectures of 1827.* R. F. Brown, P. C. Hodgson, and J. M. Stewart, trans. Berkeley: California UP, 1988. xvii, 552 p. Paper $14.95.

HÜBSCHER, ARTHUR, ed.: *Schopenhauer: Manuscript Remains: vol. I: Early Manuscripts.* New York: St. Martin's, 1988. xv, 544 p. Cloth $80.00.

INWOOD, BRAD, and L. P. GERSON, trans.: *Hellenistic Philosophy: Introductory Readings.* Indianapolis: Hackett, 1988. xii, 266 p. Cloth $26.50, paper $6.95.

KALTON, MICHAEL C., ed. and trans.: *To Become a Sage: The Ten Diagrams on Sage Learning by Yi T'oegye.* New York: Columbia, 1988. xviii, 278 p. Cloth $32.50.

LEIBNIZ, G. W.: *Discourse on Metaphysics and Related Writings.* Manchester, England: University Press, 1988. ix, 170 p. Cloth $39.95.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

TUDOR, H., and TUDOR, J. M., eds. and trans.: *Marxism and Social Democracy: The Revisionist Debate: 1896–1898.* New York: Cambridge, 1988. xi, 384 p. Cloth $49.50.

VERGOTE, ANTOINE: *Guilt and Desire: Religious Attitudes and their Pathological Derivatives.* M. H. Wood, trans. New Haven: Yale, 1988. x, 254 p. Cloth $32.50.

VICO, GIAMBATTISTA: *On the Most Ancient Wisdom of the Italians: Unearthed from the Origins of the Latin Language, Including the Disputation with the Giornale de'Letterati d'Italia.* L. M. Palmer, trans. Ithaca: Cornell, 1988. xi, 198 p. Cloth $26.95, paper $8.95.

KRIEGER BRINGS YOU A BRAND NEW TITLE IN PHILOSOPHY !!

THINKING THROUGH DEATH
VOLUMES I & II
by Scott Kramer & Kuang-ming Wu
Orig. Ed. 1988
Vol. I - 184 pp. Cl. $19.50 Ppb. $15.00
Cl. ISBN 0-89464-220-0 Ppb. ISBN 0-89464-337-1
Vol. II - 272 pp. Cl. $26.00 Ppb. $20.00
Cl. ISBN 0-89464-294-4 Ppb. ISBN 0-89464-338-X

This two volume anthology brings together, for the first time, major philosophers as they explore the meaning of death. This work contains the most representative, diverse, penetrating and readable thoughts on death ever collected. Volume I includes writings from ancient Greek and Oriental tradition to 1600. Volume II begins at the 17th century and continues to the present day. Authors include Plato, Aristotle, Confucius, Descartes, Leibniz, Nietzche, Camus, Sartre, and many others.

When ordering please add $4.00 for first book, $1.00 each additional to cover shipping. For Paperback (Ppb.) please add $2.00 first book, $.75 each additional.

KRIEGER PUBLISHING CO., INC.
P.O. Box 9542 · Melbourne, FL 32902-9542
(407) 724-9542 · Direct Order Line (407) 727-7270

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVI, NO. 4, APRIL 1989

THE BACKWARD INDUCTION PARADOX*

SUPPOSE that you and I face and know that we face a sequence of prisoner's dilemmas of known finite length: say n dilemmas. There is a well-known argument—the backward induction argument—to the effect that, in such a sequence, agents who are rational and who share the belief that they are rational will defect in every round. This argument holds however large n may be. And yet, if n is a large number, it appears that I might do better to follow a strategy such as tit-for-tat, which signals to you that I am willing to cooperate provided you reciprocate. This is the backward induction paradox.

Although game theorists have been convinced that permanent defection is the rational strategy in such a situation, they have recognized its intuitive implausibility and have often been reluctant to recommend it as a practical course of action. We believe that their hesitation is well-founded, for we hold that the argument for permanent defection is unsound and that the backward induction paradox is soluble.

I. THE PARADOX

The argument involved in the generation of the paradox involves a familiar sort of backward induction. Suppose that two players A and B face and know they face a finite sequence of n prisoner's dilemmas. Suppose also that they are both rational and that their rationality is a matter of common belief: each believes each is rational, each believes each believes this, and so on. Under those assumptions, it seems that either is in a position to run the following induction:

> My partner, being rational, will defect in the nth round of the sequence, since defecting at that stage will not have any undesirable effects in further rounds—there are none—and since it will dominate coopera-

* This paper was written while Sugden was a Visiting Fellow at the Research School of Social Science, Australian National Univ. We are grateful for a helpful discussion when it was presented at a seminar in the Department of Philosophy.

0022-362X/89/8604/169–182 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tion, just as in the one-shot prisoner's dilemma. Since he believes that I am rational, my partner will also expect me to do likewise in the nth round and so, being rational, will himself defect in round $n - 1$: there will be no undesirable effects in further rounds—the nth result is already fixed—and defecting will dominate cooperation. But, believing that I believe that he is rational, my partner will also expect me to do likewise in round $n - 1$ and so, being rational, will himself defect in round $n - 2, \ldots$.

The backward induction argument is that such an induction will lead each to the conclusion that his partner will defect in every round and that he therefore ought to do the same. If the argument is sound, then in this situation *a rational player will never cooperate.*

But can this be right? Two players who always defect will [illegible] in every round than if they had cooperated; so it is in both the interests to arrive at some cooperative understanding. Of course, this is also true of the one-shot prisoner's dilemma. But in the one-shot game the players have no means of communicating before they make their moves; and even if they *were* able to reach an agreement before the game they would have no means of enforcing it, for once a player has made a move it is too late for his partner to respond in any way. In the repeated game, in contrast, the players can communicate during the course of play by means of their choice of moves, and, in every round except the last, each player has to take account of the possibility that his choice of move may influence the later moves of his partner. Intuition surely suggests that, at least if the number of rounds is sufficiently long, a player would do better by following a conditional strategy such as tit-for-tat (that is, the strategy whereby a player cooperates in the first round and then does whatever his partner did in the previous round) than by permanently defecting. By playing such a strategy, a player signals his willingness to cooperate provided his partner will do the same, and if the signal is read correctly, it will be in the partner's interest to cooperate—in the case of tit-for-tat, in every round except the last. And even if the signal is misunderstood, little is lost.[1]

Game theorists often concede that reasonable (if not perhaps rational) people would be better advised to play conditional strategies than permanently to defect. Reinhard Selten,[2] for example, writes that "the [backward] induction theory is the game theoretically correct one but the co-operation theory [i.e. the recommendation to play conditional strategies] seems to be the better guide to practical

[1] The robustness of tit-for-tat—that is, its ability to do well against most of the strategies that a partner might plausibly play—is explained by Robert Axelrod, *The Evolution of Cooperation* (New York: Basic Books, 1984). See also Philip Pettit, "Free Riding and Foul Dealing," this JOURNAL, LXXXIII, 7 (July 1986):361–379.

[2] "The Chain Store Paradox," *Theory and Decision*, IX, 2 (1978):127–159.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

behaviour" (*ibid.*, p. 138). Similarly, Duncan Luce and Howard Raiffa,[3] while endorsing the theoretical validity of the argument for permanent defection, remark that they do not think it "reasonable" to single out this solution to the game. "By reasonable," they write, "we mean that we predict that intelligent players will play accordingly and, furthermore, that they will still do so after a full airing of the 'theory'."

The conclusion that rational players must defect seems logically inescapable, but at the same time it is intuitively implausible. That is the backward induction paradox.

II. THE SOLUTION

The solution to this paradox is surprisingly straightforward, paralleling the solution offered by Frank Jackson[4] to a version of the surprise examination paradox. The solution consists in recognizing that neither player is in a position to run the backward induction required.[5]

On careful exposition, the induction depends on the capacity of each of the players, before making his first move, to run the following sequence of arguments:

Argument 1. In round n, my partner will act rationally; therefore, he will defect.

Argument 2. In round $n - 1$, my partner will act rationally. In addition, he will then believe that I will act rationally in round n; therefore, he will believe that I will defect in round n. Given this belief, he will defect in round $n - 1$.

Argument 3. In round $n - 2$, my partner will act rationally. In addition, he will then believe, not just what he believes at round $n - 1$, but also that in round $n - 1$: (i) I will act rationally and (ii) I will believe that he will act rationally in round n. Therefore, in round $n - 2$, my partner will believe that I will defect in round $n - 1$. Given these beliefs, he will defect in round $n - 2$.

And so on, up to argument n. Taken together, these arguments yield the conclusion that the partner will defect in every round; thus, the only rational response is to defect in every round too.

But it is mistaken to think that a player is in a position to run these arguments before making his first move. At this point, he believes

[3] *Games and Decisions* (New York: Wiley, 1957), pp. 97–102.

[4] See his *Conditionals* (New York: Blackwell, 1987), ch. 7. See also Crispin Wright and Adrian Sudbury, "The Paradox of the Unexpected Examination," *Australasian Journal of Philosophy*, LV, 1 (1977):41–58.

[5] After completing this paper we discovered a rather similar critique of backward induction, but in the context of a different game, in Ken Binmore, "Modeling Rational Players," *Economics and Philosophy*, III, 2 (1987):179–214.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

that his opponent is rational, he believes that his opponent believes he is rational, and so on; this we have already assumed. But that does not entitle him to believe that in subsequent rounds his partner will still believe he is rational (etc.), irrespective of how he, the first player, has acted in the interim. Thus, for example, argument 2 is unavailable to him, because from the premise that my partner *now* believes that I am rational, I cannot derive the premise required, that in round $n - 1$ my partner will believe I am rational. Similarly, argument 3 is unavailable, because from the premise that my partner *now* believes that I *now* believe he is rational, I cannot derive the extra premise required there, that in round $n - 2$ my partner will believe that in round $n - 1$ I will believe that he is rational; and so on.

To put this point more generally, let p be any proposition that, if true at time t_1, is also true at the later time t_2. Let E be an event which may occur between t_1 and t_2. Suppose that at t_1 I believe that p. Then at t_1 must I also believe that, regardless of whether E occurs or not, I will still believe that p at t_2? Not if E is an event that I believe will not occur. If at t_1 I believe that E will not occur, then at t_1 I may believe that, were E to occur in the meantime, I would not believe p at t_2. On 31 December I may believe that it will not snow in Queensland in February. But on 31 December I may also believe that, were there to be heavy snowfalls in Queensland in late January, on 31 January I would not believe that it would not snow in Queensland in February. There is no contradiction here, provided that on 31 December I do believe that there will not be heavy snowfalls in late January.

The backward induction is supposed to show each player, at the start of the game, that rationality requires him to defect in every round—and in particular, in the first. It would indeed be irrational for me to cooperate in the first round if, regardless of whether I cooperate in that round, my partner will defect in rounds 2, . . . , n. But does the backward induction show this? Only if arguments 1, . . . , $n - 1$ all hold even on the supposition that I cooperate in round 1.

Argument 2 assumes that in round $n - 1$ my partner will believe I will act rationally in round n. But the only premise we are allowed is that at the start of the game my partner believes I am rational. Suppose that at the start of the game my partner also believes that I will not cooperate in round 1. Then we cannot argue that, even were I to cooperate in round 1, still my partner will believe in round $n - 1$ that I am rational. And similarly for arguments 3, . . . , n.

We have shown then that, just given the common belief in their rationality, neither of the players will necessarily believe that the common belief will endure. Thus, neither will necessarily be in a

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

position to run the backward induction. But the considerations rehearsed in showing this also enable us, with some supplement, to show something stronger: not just that neither will necessarily believe in the survival, come what may, of the common belief, but necessarily that neither will believe in this.

Consider the two players at the start of the game, holding a common belief in their rationality. We know that they cannot run the backward induction unless they believe that that common belief will survive, regardless of what either does.[6] But we can readily show that neither can believe that the common belief will survive, come what may; each must recognize that the belief will break down in the event of one of them cooperating.

We can show this, if we can show that any act of cooperation would cause the common belief in rationality to break down. For if *we* can show this, arguing just from the belief that the players are rational and have a common belief in their rationality, then each of the players will be in a position to do so too. Each of them will be in a position to see that, were either of them to cooperate, then the common belief in their rationality with which they start would break down.

We can show that any act of cooperation would cause the common belief in rationality to break down by starting with the nth round and seeing that cooperation would cause a breakdown there; then going back to round $n-1$ and seeing that cooperation would have a similar effect in that round; and so on, making our way back to round 1.

Suppose one of the players, say A, cooperates in the final round. This is an irrational act, since defection dominates cooperation there, and so would cause B, if rational, to believe that A is irrational. In that case, the common belief in rationality breaks down at the first level.

Suppose instead that A cooperates in round $n-1$. B may again be led to believe that A is irrational, in which case we also get a first-level breakdown of the common belief. But may B hold that A is rational? Perhaps, but only if B believes that A believes that B can be induced by A's cooperation in round $n-1$ to cooperate in round n, which would be an irrational act. So in that case the common belief breaks down too, though now at the second level up: B does not believe that A is irrational but he believes that A believes that he, B, is irrational.

Now suppose that A cooperates in round $n-2$. B may be led to believe that A is irrational. Alternatively, B may believe that A be-

[6] We do not mean to suggest that it would be sufficient to allow the players to run the backward induction if they began with the common belief and with the belief that it would survive; we think other beliefs would also be necessary, such as the common belief that it would survive, the common belief that the common belief it would survive would survive, and so on into nightmarish possibilities.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

lieves that by cooperating in round $n-2$ he can induce B to cooperate in at least one of rounds $n-1$ or n. But cooperating in round n is irrational, so this belief would involve a breakdown in the common belief at level 2. And cooperating in round $n-1$ is either irrational —in which case we also get a breakdown at level 2—or it springs from the belief that such cooperation would induce A irrationally to cooperate in round n, in which case we get a breakdown of the common belief at level 3.

The generalization should be obvious. For any act of cooperation by one player in round $n-j$, where $0 \leq j \leq n-1$, the partner, if rational, must respond with a belief that causes the common belief in rationality to break down at level $j+1$ or at some lower level. If we can see this, so can the players (who, by assumption, are rational). So, necessarily, neither of the players can believe that the common belief in rationality will survive whatever moves the players make.

We have now solved the backward induction paradox, for we have shown that the players are not necessarily in a position, and indeed are necessarily not in a position, to run the backward induction. One of the beliefs required for them to run the induction—the belief that the common belief in rationality would survive even if cooperative moves were played—is a belief that the players could not have.

III. THE SOLUTION SUPPORTED

The backward induction argument has been supposed to show that permanent defection is the uniquely rational strategy for each player. We have shown that argument to be unsound, and so we have resolved the paradox arising from the conflict between the conclusion of this argument and a more intuitive conclusion: that it could be rational for a player to cooperate in the first round, in the expectation of establishing some kind of cooperative understanding with his partner.

We have not yet shown, however, that this intuitive conclusion is correct, and we turn now to that task. We can show that it is correct, if we can show that there are beliefs which a player might hold consistently with the common belief in rationality—as we shall say, beliefs which he might rationally hold—which would make it rational for him to cooperate in the first round.

Intuitively, we might expect a player to do quite well by playing tit-for-tat, and in particular that tit-for-tat would be more successful than permanent defection. This does not mean, however, that it can be rational to play tit-for-tat. On the contrary, tit-for-tat cannot be a rational choice, since it involves cooperating without possible benefit in the last round: and this, indeed, in the expectation that the other, being rational (if this is still believed at that stage), will defect in that round.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Still, tit-for-tat points us in the right direction. From A's point of view, the intuitive argument for playing tit-for-tat is that by doing so he can induce B to believe that this *is* what he is doing. Suppose, to keep things simple, that B forms this belief as soon as he observes A's first cooperative move. Then, provided the sequence of rounds is sufficiently long, B will do best to respond in a way that induces A to cooperate from round 3 on. (If, as we suppose, B defected in round 1, there is no way he can prevent A from defecting in round 2.) One possible response is "delayed tit-for-tat": to cooperate in rounds 2 and 3 and from then on to repeat A's previous move. But this cannot be a rational choice for B, since he can gain nothing by cooperating in the last round. A better strategy is "delayed tit-for-tat minus 1," which is exactly like delayed tit-for-tat except that it always defects in the last round. If, after observing A's opening move, B forms the belief that A is playing tit-for-tat, delayed tit-for-tat minus 1 is a rational strategy: in terms of his beliefs, B can do no better than this.

Now suppose that A believes that these *are* the beliefs which B will form, and that this *is* the strategy which B will play. Then it is not rational for A to play tit-for-tat, but it *is* rational for him to play "tit-for-tat minus 2"—the strategy that is exactly like tit-for-tat except that it always defects in the last two rounds. By playing this, he makes the best possible reply to the behavior that he expects of his partner.

So, we have shown that there are beliefs which A might hold which would make it rational for him to cooperate in the first round. But are these beliefs ones which he can rationally hold? At first sight, it is tempting to think not, because A is attributing to B the belief that A is playing an irrational strategy, namely, tit-for-tat. But this is quite consistent with an initial common belief in rationality. According to A's belief, it is only after A has cooperated that B forms the belief that A is irrational. In other words, A believes that his act of cooperation will cause the common belief in rationality to break down. And this in itself cannot be used as an argument against the rationality of A's beliefs. Quite the contrary: as we showed in section II, a rational player *must* believe that any act of cooperation will cause the common belief in rationality to break down.

This is enough to show that there are beliefs which a player might rationally hold which would lead him to cooperate initially. But it is worth remarking that there are also many other beliefs which he might rationally hold which would lead him to do the same thing. Thus, just to stick with the tit-for-tat family of responses, there are different beliefs, all compatible with the initial common belief in rationality, which would rationalize different members of that family.

There is a pattern in the cases involved and we can catch it nicely in

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tabular form. The table has the advantage that it gives an idea of how the progression goes.[7]

A rationally chooses to play:	because A believes that in response to an initial cooperation B would reason in this way:
1. Tit-for-tat minus 2	A is irrational and will play tit-for-tat to the end, so I should play delayed tit-for-tat minus 1.
2. Tit-for-tat minus 4	A is rational and believes I am rational. But he believes that I believe he is irrational and that I will play delayed tit-for-tat minus 1. Hence, he will play tit-for-tat minus 2. And so I should play delayed tit-for-tat minus 3.
3. Tit-for-tat minus 6	A is rational and believes I am rational. Furthermore, A believes that I believe he is rational. But he believes that I believe that he believes that I believe that he is irrational and that I will play delayed tit-for-tat minus 3. Hence, he will play tit-for-tat minus 4. And so I should play delayed tit-for-tat minus 5.
And so on.	And so on.

Any one of the strategies listed in the left-hand column may rationally be chosen by A: each is a rational response to the corresponding beliefs held by A about B. In each case, the choice A makes involves an initial cooperation; and the belief about B that makes this a rational choice entails that the common belief in rationality breaks down. In the case in which A plays tit-for-tat minus 2, the belief breaks down at the second level: after the first round, A believes that

[7] The progression cannot go on indefinitely because the number of rounds is finite. The rationale for A's playing tit-for-tat minus j (where j is any positive even number no greater than $n - 2$) requires that this is a better response than permanent defection to a player who will respond to an initial cooperative move by playing delayed tit-for-tat minus $j - 1$. It also requires the stronger condition that delayed tit-for-tat minus $j - 1$ must be a better response than permanent defection to tit-for-tat minus $j - 2$. Let the payoffs in utility units for each player be c if they both cooperate, d if they both defect, 1 to a player who defects while his partner cooperates, and 0 to one who cooperates while his partner defects. For the classic repeated prisoner's dilemma game we require that $1 > c > d > 0$ and $c > \frac{1}{2}$. In order for tit-for-tat minus j to be rationalized, we need that $n > j + (c + d - 1)/(c - d)$. Thus, the higher the value of j, the longer the game must be in order for tit-for-tat minus j to be rationalized.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

B believes that A is irrational. In the case in which A plays tit-for-tat minus 4, it breaks down at the fourth level: A believes that B believes that A believes that B believes that A is irrational. In general, the beliefs that make it rational for A to play tit-for-tat minus j involve a breakdown in the common belief in rationality at level j.

We have shown how it is that permanently defecting is not a uniquely rational strategy and, in particular, how it is that defecting in the first round is not uniquely rational. But that leaves the question of whether it could ever be rational to defect in round 1. Could it be that initial cooperation is uniquely rational?

It had better not be, for this reason. If initial cooperation were uniquely rational, then each would believe that the other would cooperate in round 1. But this ought to have the same effect as the later belief that he cooperates or cooperated in that round. It ought to mean that the common belief in rationality does not obtain. And that is inconsistent with the assumption that rationality is a matter of common belief at the beginning.

But, in any case, it ought to be clear that it is not uniquely rational to cooperate. Consistently with the common belief in rationality a player might believe that were he to cooperate initially then the other would believe that he was irrational and beyond the strategic influence of anything the other might do. In that case, he would expect the other to defect in response to an initial cooperation and so it would be rational for him to defect in round 1.

The import of this section then is twofold. It is not uniquely rational to defect permanently, in particular, to defect in the first round. And neither is it uniquely rational to cooperate in the first round. Both options are left open by the different belief sets that rational players might hold.

IV. THE PARADOX, WITH EVERY BELIEF COMMON

We have solved the paradox presented at the beginning of this paper and we have provided some useful support for the solution. But there are related paradoxes that may be thought to create equal problems. We consider a first here and a second in the next section.

One possible objection to the argument of section III is that we have allowed A to believe that he can outguess B. In the simplest case, A plays tit-for-tat minus 2, believing that B will play delayed tit-for-tat minus 1. But if B believed that this was what A was doing, B would not play delayed tit-for-tat minus 1. (He would do better to play delayed tit-for-tat minus 3.) If A believes B to be just as rational as he is, how can he expect to outguess him?

The idea that rational players cannot expect to outguess one another is often presented in game theory as the following requirement: any belief that one player holds, or at least any belief that he

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

holds about the other, must be a common belief.[8] To see the point of this, suppose that *A* holds the beliefs *Z* that make a particular strategy *S* rational for him. Then, if all beliefs are common, *A* must believe that *B* believes that *A* believes *Z*. Thus, if *A*'s beliefs are such that *S* is rational for him, he must believe that *B* believes this to be the case too.

The idea that all beliefs must be common is thought to be a consequence of there being a common belief in rationality. Suppose *A* is rational and believes *Z*. Suppose also that *B* is rational and believes *A* to be rational. Then any reasoning that *A* can follow through can be replicated by *B*. In forming his beliefs about *B*, what has *A* to go on except his knowledge of the structure of the game and the common belief in rationality? And these are exactly the same data as *B* must use in forming his beliefs about *A*'s beliefs about *B*. How, it is asked, can rational people, reasoning from exactly the same data, arrive at different conclusions? And hence the conclusion that *B* must believe that *A* believes *Z*; and so on: that *A* believes *Z* must be a common belief.

We do not think this requirement is plausible, because it seems to us that what one player should believe about another may be underdetermined by the data at his disposal. Suppose the data available to *A* are compatible with more than one system of beliefs, and that *A* happens to light on one of these systems. There seems to be no reason why *B* should guess that that is what he comes to believe, let alone that the belief should be common between them.

But let us put these misgivings aside and assume not just that the players have a common belief in their rationality but that any belief they have about one another is also common. May the backward induction argument be invoked under those assumptions? Game theorists have thought so. But the conclusion of the argument, that rational players ought never to cooperate, offends against intuition in this case too. And so we have another version of the backward induction paradox.

Does our solution of section II serve also to solve this paradox? The answer, happily, is that it does. Even under the strengthened assumptions the player who runs the backward induction has to believe that, given at the beginning a common belief in the rationality of the players, each will continue to maintain that belief, regardless of what the other does. But he cannot believe this, for the considerations invoked in section II apply here also. He cannot believe that the common belief in rationality will survive regardless of what hap-

[8] See, e.g., R. J. Aumann, "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica*, LV, 1 (1987):1–18; or Ken Binmore and Partha Dasgupta, *Economic Organizations as Games* (New York: Blackwell, 1985), pp. 1–6.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

pens, because any act of cooperation would cause that belief to break down at some level. There is no paradox here because, as before, the argument that generates it is unsound.

It remains to provide support for that solution of the paradox by showing that the intuitive conclusion, that it can be rational to cooperate in the first round, is correct. To show this we must show that there are beliefs which a player might rationally hold which would make it rational for him to cooperate initially; and that these beliefs could be common in the sense required by the claim that a common belief in rationality entails the commonality of all beliefs.

This requires us to look more carefully at that claim. Suppose that *A* believes *Z*. The claim that therefore *B* must believe that *A* believes *Z* depends on the additional premise that *B* is rational, for if *B* were irrational there would be no grounds for expecting him to be capable of replicating whatever reasoning led *A* to his belief. The further claim that therefore *A* must believe that *B* believes that *A* believes *Z* depends on the further premises that *A* is rational and that *A* believes *B* to be rational; and so on: higher orders of belief in *Z* require higher orders of belief in the rationality of the players. The upshot of this is that, if the common belief in rationality breaks down at any level, so too does the argument for the commonality of all beliefs.

At the beginning of the sequence, when the common belief in rationality is in place, the argument for the commonality of all beliefs goes through, at least given our earlier concession. But since any act of cooperation will undermine the common belief in rationality at some level, we know that equally it must undermine the case for the commonality of whatever beliefs are held by the players.

With this point in mind, let us now ask whether it could be rational under the strengthened assumptions for either of the players to cooperate in the first round as part of one of the strategies discussed in the last section: tit-for-tat minus 2, tit-for-tat minus 4, and so on. If it could be rational to do this, then the support provided for the solution in the case of the original paradox serves also to support the solution here.

Consider the simplest of the patterns of beliefs we presented in section III—the pattern of beliefs which leads *A* to play tit-for-tat minus 2. *A* believes that *B* has the initial belief that *A* will defect in every round. *A* also believes, however, that, were he to cooperate in round 1, *B* would be led to believe that he was irrationally playing tit-for-tat and would therefore respond by playing delayed tit-for-tat minus 1. In the light of all these beliefs, *A* plays tit-for-tat minus 2. To this it might be objected that *B*, as a rational player, ought to be capable of replicating *A*'s reasoning and hence of working out what *A*'s beliefs are; and that, if he did this, he would realize that *A* was

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

playing tit-for-tat minus 2 and not tit-for-tat. Furthermore (the objection would go), *A* ought to be able to see that *B* would realize this. Is there therefore not an inconsistency in *A*'s beliefs? But by now it should be clear that the answer has to be "no." *A* believes that *B* believes he is irrational, so in *A*'s belief, *B* has no reason to replicate the reasoning through which he, *A*, has in fact gone. Thus, *A* may rationally hold the beliefs that would lead him to play tit-for-tat minus 2 even under the assumption that each player's initial beliefs about his partner must be matters of common belief.

It should be clear that this defense of the possible rationality of playing tit-for-tat minus 2 can be replicated in defense of the possible rationality of tit-for-tat minus 4, and so on. We may conclude then that, even if a common belief in rationality entails the commonality of all beliefs, still it may be rational for a player to cooperate initially.

This means that, contrary to the conclusion of the backward induction argument, the strategy of permanent defection, and in particular any strategy involving initial defection, is not uniquely rational. But, by the same argument as that presented in the last section, we can see that neither is initial cooperation uniquely rational: a player would be rational to defect initially if he believed, as he still might, that an initial cooperation would cause the other player to believe that he was irrational and beyond the reach of strategic considerations.

We must conclude, then, that the solution provided for our original paradox, and the support given to that solution, are both still relevant in the case of the strengthened paradox. The assumption that the common belief in rationality entails the commonality of all beliefs, even if it is conceded to be reasonable, does not significantly complicate things.[9]

V. THE PARADOX, WITH COMMON KNOWLEDGE OF RATIONALITY

Some game theorists postulate, not just that the players have a common belief in their rationality, but that this rationality is a matter of common knowledge: it is not something on which they could be mistaken, could think they were mistaken, could think they thought they were mistaken, and so on. What becomes of our paradox and our solution under this way of strengthening the assumptions?

It should be clear that, whatever is true about the paradox, the solution fails. Under the assumption of common knowledge of their rationality, each of the players is able rationally to believe that the common belief in rationality will endure; indeed, each knows that common knowledge will endure: otherwise, he could not know it to be knowledge. Thus, each is in a position where he can rationally

[9] Under this assumption, many game theorists will wish to argue that permanent defection is a sequential equilibrium. A critique of the significance of that argument is developed by Binmore in "Modeling Rational Players."

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

endorse arguments 1 . . . n of section II, or versions of these arguments with 'know' and 'knowledge' substituted for 'believe' and 'belief'. Each can run the backward induction and each will therefore defect.

What we wish to argue, however, is that, if the solution to our paradox fails under common knowledge of rationality, so does the paradox itself. The paradox arises from the conflict between the backward induction argument for permanent defection and the more intuitive thought that a player might do better to play a conditional strategy involving an initial act of cooperation. But this intuitive thought depends on a particular way of thinking about games which, although entirely natural, is not allowed by the assumption of common knowledge. The intuitive idea is that each player is free to choose any one of the strategies that he is allowed by the rules of the game. To decide which strategy to choose, the player asks himself what would happen in the event that he chose each strategy. If he is rational, he must then choose the strategy that leads to the best consequences; but it is possible for him to discover that this is the best strategy only because the hypothesis that he chooses a different (and, as it turns out, irrational) strategy is intelligible. Confronted with the backward induction argument for defection, the intuitive response is to think that, despite the apparent logic of that argument, the consequences following from a conditional strategy might be preferable to those following from permanent defection.

But if the players have common knowledge of their rationality, this intuitive response misses the point, for in a sense the backward induction argument does not show that a player does better by defecting: it shows that, as a matter of logical necessity, both players *must* defect and presumably therefore that they know they must defect. The reason is that, when a player A makes the intuitive response and asks himself how B is likely to respond in the event of A's initially cooperating, he is asking after what would happen under an inconsistent hypothesis. For it is inconsistent to postulate both that there is common knowledge of rationality, in which case there is common knowledge that both will defect in every round, and that one of the players cooperates in round 1.

If this analysis seems too quick, then a parallel may help to make the point vivid. I may believe that of two options before me, p and q, p promises the better outcome, and still ask what would happen, the world being as it is, were I to choose q instead; indeed, I am very likely to do this in order to check my belief that p does actually hold out the better prospect. I am able to ask this question, because there is nothing inconsistent in the hypothesis that, the world being as it is, I might believe that q was better and choose it rather than p. But suppose it is required now that knowledge states—indeed, states

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

involving knowing one knows, knowing one knows one knows, and so on—replace belief states. In that case, I cannot ask after what would happen, the world being as it is, were I to know that q is better; the world being as it is, i.e., p being actually better, it is not possible that I might know that q was better.

The situation where the players are ascribed common knowledge of their rationality ought strictly to have no interest for game theory. Under the assumption of common knowledge, neither player is allowed to think strategically. On pain of inconsistency, neither can ask after what effect he might have on the other through doing something besides what that assumption entails he should do: viz., permanently defect. We believe that the assumption of common knowledge of rationality, as it is invoked in game theory, should not be taken completely literally. It should probably be taken to mean something like what we have meant by 'common belief'.[10]

CONCLUSION

The backward induction paradox, as presented in sections I and IV, is solved by the fact that rational players are necessarily not in a position to run the backward induction argument. That is our principal claim. But two subsidary theses are also worth noting. The first is that, in the situations that give rise to the paradox, neither an initial cooperation nor an initial defection is uniquely rational: either may be rational in terms of beliefs that a player can rationally hold. The second is that there is no backward induction paradox under the one assumption we have considered which would license the use of the backward induction argument: the assumption of common knowledge of rationality. Understood literally, this assumption is inconsistent with what we take to be the whole point of game theory —the analysis of strategic thinking.[11]

PHILIP PETTIT

Australian National University

ROBERT SUGDEN

University of East Anglia

[10] Recent work in game theory has implicitly recognized this: see R. Selten, "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory*, IV, 1 (1975):25–55; and D. M. Kreps and R. Wilson, "Sequential Equilibrium," *Econometrica*, I., 4 (1982):863–894. Selten's concept of "perfect equilibrium" and Kreps and Wilson's concept of "sequential equilibrium" are both based on the idea that a player must initially assume that other players are rational, but that each player's strategy must include contingency plans to be brought into action if this initial assumption is disproved in the course of the game. The usual way of describing this approach is to say that it is common knowledge that the players are almost certain to be rational; this has much the same effect as making their rationality a matter of common belief.

[11] After this paper was in proof, an article by Cristina Bicchieri, in which she introduces an argument similar to that presented here, was brought to our attention [cf. "Self-refuting Theories of Strategic Interaction: A Paradox of Common Knowledge," *Erkenntnis* (forthcoming)].

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

WHY ABORTION IS IMMORAL

THE view that abortion is, with rare exceptions, seriously immoral has received little support in the recent philosophical literature. No doubt most philosophers affiliated with secular institutions of higher education believe that the anti-abortion position is either a symptom of irrational religious dogma or a conclusion generated by seriously confused philosophical argument. The purpose of this essay is to undermine this general belief. This essay sets out an argument that purports to show, as well as any argument in ethics can show, that abortion is, except possibly in rare cases, seriously immoral, that it is in the same moral category as killing an innocent adult human being.

The argument is based on a major assumption. Many of the most insightful and careful writers on the ethics of abortion—such as Joel Feinberg, Michael Tooley, Mary Anne Warren, H. Tristram Engelhardt, Jr., L. W. Sumner, John T. Noonan, Jr., and Philip Devine[1]—believe that whether or not abortion is morally permissible stands or falls on whether or not a fetus is the sort of being whose life it is seriously wrong to end. The argument of this essay will assume, but not argue, that they are correct.

Also, this essay will neglect issues of great importance to a complete ethics of abortion. Some anti-abortionists will allow that certain abortions, such as abortion before implantation or abortion when the life of a woman is threatened by a pregnancy or abortion after rape, may be morally permissible. This essay will not explore the casuistry of these hard cases. The purpose of this essay is to develop a general argument for the claim that the overwhelming majority of deliberate abortions are seriously immoral.

I.

A sketch of standard anti-abortion and pro-choice arguments exhibits how those arguments possess certain symmetries that explain why partisans of those positions are so convinced of the correctness of their own positions, why they are not successful in convincing

[1] Feinberg, "Abortion," in *Matters of Life and Death: New Introductory Essays in Moral Philosophy*, Tom Regan, ed. (New York: Random House, 1986), pp. 256–293; Tooley, "Abortion and Infanticide," *Philosophy and Public Affairs*, II, 1 (1972):37–65, Tooley, *Abortion and Infanticide* (New York: Oxford, 1984); Warren, "On the Moral and Legal Status of Abortion," *The Monist*, LVII, 1 (1973):43–61; Engelhardt, "The Ontology of Abortion," *Ethics*, LXXXIV, 3 (1974):217–234; Sumner, *Abortion and Moral Theory* (Princeton: University Press, 1981); Noonan, "An Almost Absolute Value in History," in *The Morality of Abortion: Legal and Historical Perspectives*, Noonan, ed. (Cambridge: Harvard, 1970); and Devine, *The Ethics of Homicide* (Ithaca: Cornell, 1978).

0022-362X/89/8604/183–202 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

their opponents, and why, to others, this issue seems to be unresolvable. An analysis of the nature of this standoff suggests a strategy for surmounting it.

Consider the way a typical anti-abortionist argues. She will argue or assert that life is present from the moment of conception or that fetuses look like babies or that fetuses possess a characteristic such as a genetic code that is both necessary and sufficient for being human. Anti–abortionists seem to believe that (1) the truth of all of these claims is quite obvious, and (2) establishing any of these claims is sufficient to show that abortion is morally akin to murder.

A standard pro-choice strategy exhibits similarities. The pro-choicer will argue or assert that fetuses are not persons or that fetuses are not rational agents or that fetuses are not social beings. Pro-choicers seem to believe that (1) the truth of any of these claims is quite obvious, and (2) establishing any of these claims is sufficient to show that an abortion is not a wrongful killing.

In fact, both the pro-choice and the anti-abortion claims do seem to be true, although the "it looks like a baby" claim is more difficult to establish the earlier the pregnancy. We seem to have a standoff. How can it be resolved?

As everyone who has taken a bit of logic knows, if any of these arguments concerning abortion is a good argument, it requires not only some claim characterizing fetuses, but also some general moral principle that ties a characteristic of fetuses to having or not having the right to life or to some other moral characteristic that will generate the obligation or the lack of obligation not to end the life of a fetus. Accordingly, the arguments of the anti-abortionist and the pro-choicer need a bit of filling in to be regarded as adequate.

Note what each partisan will say. The anti-abortionist will claim that her position is supported by such generally accepted moral principles as "It is always prima facie seriously wrong to take a human life" or "It is always prima facie seriously wrong to end the life of a baby." Since these are generally accepted moral principles, her position is certainly not obviously wrong. The pro-choicer will claim that her position is supported by such plausible moral principles as "Being a person is what gives an individual intrinsic moral worth" or "It is only seriously prima facie wrong to take the life of a member of the human community." Since these are generally accepted moral principles, the pro-choice position is certainly not obviously wrong. Unfortunately, we have again arrived at a standoff.

Now, how might one deal with this standoff? The standard approach is to try to show how the moral principles of one's opponent lose their plausibility under analysis. It is easy to see how this is

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

possible. On the one hand, the anti-abortionist will defend a moral principle concerning the wrongness of killing which tends to be broad in scope in order that even fetuses at an early stage of pregnancy will fall under it. The problem with broad principles is that they often embrace too much. In this particular instance, the principle "It is always prima facie wrong to take a human life" seems to entail that it is wrong to end the existence of a living human cancer-cell culture, on the grounds that the culture is both living and human. Therefore, it seems that the anti-abortionist's favored principle is too broad.

On the other hand, the pro-choicer wants to find a moral principle concerning the wrongness of killing which tends to be narrow in scope in order that fetuses will *not* fall under it. The problem with narrow principles is that they often do not embrace enough. Hence, the needed principles such as "It is prima facie seriously wrong to kill only persons" or "It is prima facie wrong to kill only rational agents" do not explain why it is wrong to kill infants or young children or the severely retarded or even perhaps the severely mentally ill. Therefore, we seem again to have a standoff. The anti-abortionist charges, not unreasonably, that pro–choice principles concerning killing are too narrow to be acceptable; the pro-choicer charges, not unreasonably, that anti-abortionist principles concerning killing are too broad to be acceptable.

Attempts by both sides to patch up the difficulties in their positions run into further difficulties. The anti-abortionist will try to remove the problem in her position by reformulating her principle concerning killing in terms of human beings. Now we end up with: "It is always prima facie seriously wrong to end the life of a human being." This principle has the advantage of avoiding the problem of the human cancer-cell culture counterexample. But this advantage is purchased at a high price. For although it is clear that a fetus is both human and alive, it is not at all clear that a fetus is a human *being*. There is at least something to be said for the view that something becomes a human being only after a process of development, and that therefore first trimester fetuses and perhaps all fetuses are not yet human beings. Hence, the anti-abortionist, by this move, has merely exchanged one problem for another.[2]

The pro-choicer fares no better. She may attempt to find reasons why killing infants, young children, and the severely retarded is

[2] For interesting discussions of this issue, see Warren Quinn, "Abortion: Identity and Loss," *Philosophy and Public Affairs*, XIII, 1 (1984):24–54; and Lawrence C. Becker, "Human Being: The Boundaries of the Concept," *Philosophy and Public Affairs*, IV, 4 (1975):334–359.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

wrong which are independent of her major principle that is supposed to explain the wrongness of taking human life, but which will not also make abortion immoral. This is no easy task. Appeals to social utility will seem satisfactory only to those who resolve not to think of the enormous difficulties with a utilitarian account of the wrongness of killing and the significant social costs of preserving the lives of the unproductive.[3] A pro-choice strategy that extends the definition of 'person' to infants or even to young children seems just as arbitrary as an anti-abortion strategy that extends the definition of 'human being' to fetuses. Again, we find symmetries in the two positions and we arrive at a standoff.

There are even further problems that reflect symmetries in the two positions. In addition to counterexample problems, or the arbitrary application problems that can be exchanged for them, the standard anti–abortionist principle "It is prima facie seriously wrong to kill a human being," or one of its variants, can be objected to on the grounds of ambiguity. If 'human being' is taken to be a *biological* category, then the anti-abortionist is left with the problem of explaining why a merely biological category should make a moral difference. Why, it is asked, is it any more reasonable to base a moral conclusion on the number of chromosomes in one's cells than on the color of one's skin?[4] If 'human being', on the other hand, is taken to be a *moral* category, then the claim that a fetus is a human being cannot be taken to be a premise in the anti–abortion argument, for it is precisely what needs to be established. Hence, either the anti-abortionist's main category is a morally irrelevant, merely biological category, or it is of no use to the anti-abortionist in establishing (noncircularly, of course) that abortion is wrong.

Although this problem with the anti-abortionist position is often noticed, it is less often noticed that the pro-choice position suffers from an analogous problem. The principle "Only persons have the right to life" also suffers from an ambiguity. The term 'person' is typically defined in terms of psychological characteristics, although there will certainly be disagreement concerning which characteristics are most important. Supposing that this matter can be settled, the pro-choicer is left with the problem of explaining why *psychological* characteristics should make a *moral* difference. If the pro-choicer should attempt to deal with this problem by claiming that an explana-

[3] For example, see my "Ethics and The Elderly: Some Problems," in Stuart Spicker, Kathleen Woodward, and David Van Tassel, eds., *Aging and the Elderly: Humanistic Perspectives in Gerontology* (Atlantic Highlands, NJ: Humanities, 1978), pp. 341–355.

[4] See Warren, *op. cit.*, and Tooley, "Abortion and Infanticide."

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tion is not necessary, that in fact we do treat such a cluster of psychological properties as having moral significance, the sharp-witted anti-abortionist should have a ready response. We do treat being both living and human as having moral significance. If it is legitimate for the pro-choicer to demand that the anti-abortionist provide an explanation of the connection between the biological character of being a human being and the wrongness of being killed (even though people accept this connection), then it is legitimate for the anti-abortionist to demand that the pro-choicer provide an explanation of the connection between psychological criteria for being a person and the wrongness of being killed (even though that connection is accepted).[5]

Feinberg has attempted to meet this objection (he calls psychological personhood "commonsense personhood"):

> The characteristics that confer commonsense personhood are not arbitrary bases for rights and duties, such as race, sex or species membership; rather they are traits that make sense out of rights and duties and without which those moral attributes would have no point or function. It is because people are conscious; have a sense of their personal identities; have plans, goals, and projects; experience emotions; are liable to pains, anxieties, and frustrations; can reason and bargain, and so on—it is because of these attributes that people have values and interests, desires and expectations of their own, including a stake in their own futures, and a personal well-being of a sort we cannot ascribe to unconscious or nonrational beings. Because of their developed capacities they can assume duties and responsibilities and can have and make claims on one another. Only because of their sense of self, their life plans, their value hierarchies, and their stakes in their own futures can they be ascribed fundamental rights. There is nothing arbitrary about these linkages (*op. cit.*, p. 270).

The plausible aspects of this attempt should not be taken to obscure its implausible features. There is a great deal to be said for the view that being a psychological person under some description is a necessary condition for having duties. One cannot have a duty unless one is capable of behaving morally, and a being's capability of behaving morally will require having a certain psychology. It is far from obvious, however, that having rights entails consciousness or rationality, as Feinberg suggests. We speak of the rights of the severely retarded or the severely mentally ill, yet some of these persons are not rational. We speak of the rights of the temporarily unconscious. The New Jersey Supreme Court based their decision in the Quinlan case

[5] This seems to be the fatal flaw in Warren's treatment of this issue.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

on Karen Ann Quinlan's right to privacy, and she was known to be permanently unconscious at that time. Hence, Feinberg's claim that having rights entails being conscious is, on its face, obviously false.

Of course, it might not make sense to attribute rights to a being that would never in its natural history have certain psychological traits. This modest connection between psychological personhood and moral personhood will create a place for Karen Ann Quinlan and the temporarily unconscious. But then it makes a place for fetuses also. Hence, it does not serve Feinberg's pro-choice purposes. Accordingly, it seems that the pro-choicer will have as much difficulty bridging the gap between psychological personhood and personhood in the moral sense as the anti-abortionist has bridging the gap between being a biological human being and being a human being in the moral sense.

Furthermore, the pro-choicer cannot any more escape her problem by making person a purely moral category than the anti-abortionist could escape by the analogous move. For if person is a moral category, then the pro-choicer is left without the resources for establishing (noncircularly, of course) the claim that a fetus is not a person, which is an essential premise in her argument. Again, we have both a symmetry and a standoff between pro-choice and anti-abortion views.

Passions in the abortion debate run high. There are both plausibilities and difficulties with the standard positions. Accordingly, it is hardly surprising that partisans of either side embrace with fervor the moral generalizations that support the conclusions they preanalytically favor, and reject with disdain the moral generalizations of their opponents as being subject to inescapable difficulties. It is easy to believe that the counterexamples to one's own moral principles are merely temporary difficulties that will dissolve in the wake of further philosophical research, and that the counterexamples to the principles of one's opponents are as straightforward as the contradiction between *A* and *O* propositions in traditional logic. This might suggest to an impartial observer (if there are any) that the abortion issue is unresolvable.

There is a way out of this apparent dialectical quandary. The moral generalizations of both sides are not quite correct. The generalizations hold for the most part, for the usual cases. This suggests that they are all *accidental* generalizations, that the moral claims made by those on both sides of the dispute do not touch on the *essence* of the matter.

This use of the distinction between essence and accident is not meant to invoke obscure metaphysical categories. Rather, it is in-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tended to reflect the rather atheoretical nature of the abortion discussion. If the generalization a partisan in the abortion dispute adopts were derived from the reason why ending the life of a human being is wrong, then there could not be exceptions to that generalization unless some special case obtains in which there are even more powerful countervailing reasons. Such generalizations would not be merely accidental generalizations; they would point to, or be based upon, the essence of the wrongness of killing, what it is that makes killing wrong. All this suggests that a necessary condition of resolving the abortion controversy is a more theoretical account of the wrongness of killing. After all, if we merely believe, but do not understand, why killing adult human beings such as ourselves is wrong, how could we conceivably show that abortion is either immoral or permissible?

II.

In order to develop such an account, we can start from the following unproblematic assumption concerning our own case: it is wrong to kill *us*. Why is it wrong? Some answers can be easily eliminated. It might be said that what makes killing us wrong is that a killing brutalizes the one who kills. But the brutalization consists of being inured to the performance of an act that is hideously immoral; hence, the brutalization does not explain the immorality. It might be said that what makes killing us wrong is the great loss others would experience due to our absence. Although such hubris is understandable, such an explanation does not account for the wrongness of killing hermits, or those whose lives are relatively independent and whose friends find it easy to make new friends.

A more obvious answer is better. What primarily makes killing wrong is neither its effect on the murderer nor its effect on the victim's friends and relatives, but its effect on the victim. The loss of one's life is one of the greatest losses one can suffer. The loss of one's life deprives one of all the experiences, activities, projects, and enjoyments that would otherwise have constituted one's future. Therefore, killing someone is wrong, primarily because the killing inflicts (one of) the greatest possible losses on the victim. To describe this as the loss of life can be misleading, however. The change in my biological state does not by itself make killing me wrong. The effect of the loss of my biological life is the loss to me of all those activities, projects, experiences, and enjoyments which would otherwise have constituted my future personal life. These activities, projects, experiences, and enjoyments are either valuable for their own sakes or are means to something else that is valuable for its own sake. Some parts of my future are not valued by me now, but will come to be valued by

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

me as I grow older and as my values and capacities change. When I am killed, I am deprived both of what I now value which would have been part of my future personal life, but also what I would come to value. Therefore, when I die, I am deprived of all of the value of my future. Inflicting this loss on me is ultimately what makes killing me wrong. This being the case, it would seem that what makes killing *any* adult human being prima facie seriously wrong is the loss of his or her future.[6]

How should this rudimentary theory of the wrongness of killing be evaluated? It cannot be faulted for deriving an 'ought' from an 'is', for it does not. The analysis assumes that killing me (or you, reader) is prima facie seriously wrong. The point of the analysis is to establish which natural property ultimately explains the wrongness of the killing, given that it is wrong. A natural property will ultimately explain the wrongness of killing, only if (1) the explanation fits with our intuitions about the matter and (2) there is no other natural property that provides the basis for a better explanation of the wrongness of killing. This analysis rests on the intuition that what makes killing a particular human or animal wrong is what it does to that particular human or animal. What makes killing wrong is some natural effect or other of the killing. Some would deny this. For instance, a divine-command theorist in ethics would deny it. Surely this denial is, however, one of those features of divine–command theory which renders it so implausible.

The claim that what makes killing wrong is the loss of the victim's future is directly supported by two considerations. In the first place, this theory explains why we regard killing as one of the worst of crimes. Killing is especially wrong, because it deprives the victim of more than perhaps any other crime. In the second place, people with AIDS or cancer who know they are dying believe, of course, that dying is a very bad thing for them. They believe that the loss of a future to them that they would otherwise have experienced is what makes their premature death a very bad thing for them. A better theory of the wrongness of killing would require a different natural property associated with killing which better fits with the attitudes of the dying. What could it be?

The view that what makes killing wrong is the loss to the victim of the value of the victim's future gains additional support when some of its implications are examined. In the first place, it is incompatible

[6] I have been most influenced on this matter by Jonathan Glover, *Causing Death and Saving Lives* (New York: Penguin, 1977), ch. 3; and Robert Young, "What Is So Wrong with Killing People?" *Philosophy*, LIV, 210 (1979):515–528.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

with the view that it is wrong to kill only beings who are biologically human. It is possible that there exists a different species from another planet whose members have a future like ours. Since having a future like that is what makes killing someone wrong, this theory entails that it would be wrong to kill members of such a species. Hence, this theory is opposed to the claim that only life that is biologically human has great moral worth, a claim which many anti-abortionists have seemed to adopt. This opposition, which this theory has in common with personhood theories, seems to be a merit of the theory.

In the second place, the claim that the loss of one's future is the wrong-making feature of one's being killed entails the possibility that the futures of some actual nonhuman mammals on our own planet are sufficiently like ours that it is seriously wrong to kill them also. Whether some animals do have the same right to life as human beings depends on adding to the account of the wrongness of killing some additional account of just what it is about my future or the futures of other adult human beings which makes it wrong to kill us. No such additional account will be offered in this essay. Undoubtedly, the provision of such an account would be a very difficult matter. Undoubtedly, any such account would be quite controversial. Hence, it surely should not reflect badly on this sketch of an elementary theory of the wrongness of killing that it is indeterminate with respect to some very difficult issues regarding animal rights.

In the third place, the claim that the loss of one's future is the wrong-making feature of one's being killed does not entail, as sanctity of human life theories do, that active euthanasia is wrong. Persons who are severely and incurably ill, who face a future of pain and despair, and who wish to die will not have suffered a loss if they are killed. It is, strictly speaking, the value of a human's future which makes killing wrong in this theory. This being so, killing does not necessarily wrong some persons who are sick and dying. Of course, there may be other reasons for a prohibition of active euthanasia, but that is another matter. Sanctity-of-human-life theories seem to hold that active euthanasia is seriously wrong even in an individual case where there seems to be good reason for it independently of public policy considerations. This consequence is most implausible, and it is a plus for the claim that the loss of a future of value is what makes killing wrong that it does not share this consequence.

In the fourth place, the account of the wrongness of killing defended in this essay does straightforwardly entail that it is prima facie seriously wrong to kill children and infants, for we do presume that they have futures of value. Since we do believe that it is wrong to kill

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

defenseless little babies, it is important that a theory of the wrongness of killing easily account for this. Personhood theories of the wrongness of killing, on the other hand, cannot straightforwardly account for the wrongness of killing infants and young children.[7] Hence, such theories must add special ad hoc accounts of the wrongness of killing the young. The plausibility of such ad hoc theories seems to be a function of how desperately one wants such theories to work. The claim that the primary wrong-making feature of a killing is the loss to the victim of the value of its future accounts for the wrongness of killing young children and infants directly; it makes the wrongness of such acts as obvious as we actually think it is. This is a further merit of this theory. Accordingly, it seems that this value of a future-like-ours theory of the wrongness of killing shares strengths of both sanctity-of-life and personhood accounts while avoiding weaknesses of both. In addition, it meshes with a central intuition concerning what makes killing wrong.

The claim that the primary wrong-making feature of a killing is the loss to the victim of the value of its future has obvious consequences for the ethics of abortion. The future of a standard fetus includes a set of experiences, projects, activities, and such which are identical with the futures of adult human beings and are identical with the futures of young children. Since the reason that is sufficient to explain why it is wrong to kill human beings after the time of birth is a reason that also applies to fetuses, it follows that abortion is prima facie seriously morally wrong.

This argument does not rely on the invalid inference that, since it is wrong to kill persons, it is wrong to kill potential persons also. The category that is morally central to this analysis is the category of having a valuable future like ours; it is not the category of personhood. The argument to the conclusion that abortion is prima facie seriously morally wrong proceeded independently of the notion of person or potential person or any equivalent. Someone may wish to start with this analysis in terms of the value of a human future, conclude that abortion is, except perhaps in rare circumstances, seriously morally wrong, infer that fetuses have the right to life, and then call fetuses "persons" as a result of their having the right to life. Clearly, in this case, the category of person is being used to state the *conclusion* of the analysis rather than to generate the *argument* of the analysis.

The structure of this anti-abortion argument can be both illuminated and defended by comparing it to what appears to be the best

[7] Feinberg, Tooley, Warren, and Engelhardt have all dealt with this problem.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

argument for the wrongness of the wanton infliction of pain on animals. This latter argument is based on the assumption that it is prima facie wrong to inflict pain on me (or you, reader). What is the natural property associated with the infliction of pain which makes such infliction wrong? The obvious answer seems to be that the infliction of pain causes suffering and that suffering is a misfortune. The suffering caused by the infliction of pain is what makes the wanton infliction of pain on me wrong. The wanton infliction of pain on other adult humans causes suffering. The wanton infliction of pain on animals causes suffering. Since causing suffering is what makes the wanton infliction of pain wrong and since the wanton infliction of pain on animals causes suffering, it follows that the wanton infliction of pain on animals is wrong.

This argument for the wrongness of the wanton infliction of pain on animals shares a number of structural features with the argument for the serious prima facie wrongness of abortion. Both arguments start with an obvious assumption concerning what it is wrong to do to me (or you, reader). Both then look for the characteristic or the consequence of the wrong action which makes the action wrong. Both recognize that the wrong–making feature of these immoral actions is a property of actions sometimes directed at individuals other than postnatal human beings. If the structure of the argument for the wrongness of the wanton infliction of pain on animals is sound, then the structure of the argument for the prima facie serious wrongness of abortion is also sound, for the structure of the two arguments is the same. The structure common to both is the key to the explanation of how the wrongness of abortion can be demonstrated without recourse to the category of person. In neither argument is that category crucial.

This defense of an argument for the wrongness of abortion in terms of a structurally similar argument for the wrongness of the wanton infliction of pain on animals succeeds only if the account regarding animals is the correct account. Is it? In the first place, it seems plausible. In the second place, its major competition is Kant's account. Kant believed that we do not have direct duties to animals at all, because they are not persons. Hence, Kant had to explain and justify the wrongness of inflicting pain on animals on the grounds that "he who is hard in his dealings with animals becomes hard also in his dealing with men."[8] The problem with Kant's account is that

[8] "Duties to Animals and Spirits," in *Lectures on Ethics*, Louis Infeld, trans. (New York: Harper, 1963), p. 239.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

there seems to be no reason for accepting this latter claim unless Kant's account is rejected. If the alternative to Kant's account is accepted, then it is easy to understand why someone who is indifferent to inflicting pain on animals is also indifferent to inflicting pain on humans, for one is indifferent to what makes inflicting pain wrong in both cases. But, if Kant's account is accepted, there is no intelligible reason why one who is hard in his dealings with animals (or crabgrass or stones) should also be hard in his dealings with men. After all, men are persons: animals are no more persons than crabgrass or stones. Persons are Kant's crucial moral category. Why, in short, should a Kantian accept the basic claim in Kant's argument?

Hence, Kant's argument for the wrongness of inflicting pain on animals rests on a claim that, in a world of Kantian moral agents, is demonstrably false. Therefore, the alternative analysis, being more plausible anyway, should be accepted. Since this alternative analysis has the same structure as the anti-abortion argument being defended here, we have further support for the argument for the immorality of abortion being defended in this essay.

Of course, this value of a future-like-ours argument, if sound, shows only that abortion is prima facie wrong, not that it is wrong in any and all circumstances. Since the loss of the future to a standard fetus, if killed, is, however, at least as great a loss as the loss of the future to a standard adult human being who is killed, abortion, like ordinary killing, could be justified only by the most compelling reasons. The loss of one's life is almost the greatest misfortune that can happen to one. Presumably abortion could be justified in some circumstances, only if the loss consequent on failing to abort would be at least as great. Accordingly, morally permissible abortions will be rare indeed unless, perhaps, they occur so early in pregnancy that a fetus is not yet definitely an individual. Hence, this argument should be taken as showing that abortion is presumptively very seriously wrong, where the presumption is very strong—as strong as the presumption that killing another adult human being is wrong.

III.

How complete an account of the wrongness of killing does the value of a future-like-ours account have to be in order that the wrongness of abortion is a consequence? This account does not have to be an account of the necessary conditions for the wrongness of killing. Some persons in nursing homes may lack valuable human futures, yet it may be wrong to kill them for other reasons. Furthermore, this account does not obviously have to be the sole reason killing is wrong where the victim did have a valuable future. This analysis claims only

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

that, for any killing where the victim did have a valuable future like ours, having that future by itself is sufficient to create the strong presumption that the killing is seriously wrong.

One way to overturn the value of a future-like-ours argument would be to find some account of the wrongness of killing which is at least as intelligible and which has different implications for the ethics of abortion. Two rival accounts possess at least some degree of plausibility. One account is based on the obvious fact that people value the experience of living and wish for that valuable experience to continue. Therefore, it might be said, what makes killing wrong is the discontinuation of that experience for the victim. Let us call this the *discontinuation account.*[9] Another rival account is based upon the obvious fact that people strongly desire to continue to live. This suggests that what makes killing us so wrong is that it interferes with the fulfillment of a strong and fundamental desire, the fulfillment of which is necessary for the fulfillment of any other desires we might have. Let us call this the *desire account.*[10]

Consider first the desire account as a rival account of the ethics of killing which would provide the basis for rejecting the anti-abortion position. Such an account will have to be stronger than the value of a future-like-ours account of the wrongness of abortion if it is to do the job expected of it. To entail the wrongness of abortion, the value of a future-like-ours account has only to provide a sufficient, but not a necessary, condition for the wrongness of killing. The desire account, on the other hand, must provide us also with a necessary condition for the wrongness of killing in order to generate a pro-choice conclusion on abortion. The reason for this is that presumably the argument from the desire account moves from the claim that what makes killing wrong is interference with a very strong desire to the claim that abortion is not wrong because the fetus lacks a strong desire to live. Obviously, this inference fails if someone's having the desire to live is not a necessary condition of its being wrong to kill that individual.

One problem with the desire account is that we do regard it as seriously wrong to kill persons who have little desire to live or who have no desire to live or, indeed, have a desire not to live. We believe it is seriously wrong to kill the unconscious, the sleeping, those who

[9] I am indebted to Jack Bricke for raising this objection.

[10] Presumably a preference utilitarian would press such an objection. Tooley once suggested that his account has such a theoretical underpinning. See his "Abortion and Infanticide," pp. 44/5.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

are tired of life, and those who are suicidal. The value-of-a-human-future account renders standard morality intelligible in these cases; these cases appear to be incompatible with the desire account.

The desire account is subject to a deeper difficulty. We desire life, because we value the goods of this life. The goodness of life is not secondary to our desire for it. If this were not so, the pain of one's own premature death could be done away with merely by an appropriate alteration in the configuration of one's desires. This is absurd. Hence, it would seem that it is the loss of the goods of one's future, not the interference with the fulfillment of a strong desire to live, which accounts ultimately for the wrongness of killing.

It is worth noting that, if the desire account is modified so that it does not provide a necessary, but only a sufficient, condition for the wrongness of killing, the desire account is compatible with the value of a future-like-ours account. The combined accounts will yield an anti-abortion ethic. This suggests that one can retain what is intuitively plausible about the desire account without a challenge to the basic argument of this paper.

It is also worth noting that, if future desires have moral force in a modified desire account of the wrongness of killing, one can find support for an anti-abortion ethic even in the absence of a value of a future-like-ours account. If one decides that a morally relevant property, the possession of which is sufficient to make it wrong to kill some individual, is the desire at some future time to live—one might decide to justify one's refusal to kill suicidal teenagers on these grounds, for example—then, since typical fetuses will have the desire in the future to live, it is wrong to kill typical fetuses. Accordingly, it does not seem that a desire account of the wrongness of killing can provide a justification of a pro-choice ethic of abortion which is nearly as adequate as the value of a human-future justification of an anti-abortion ethic.

The discontinuation account looks more promising as an account of the wrongness of killing. It seems just as intelligible as the value of a future-like-ours account, but it does not justify an anti-abortion position. Obviously, if it is the continuation of one's activities, experiences, and projects, the loss of which makes killing wrong, then it is not wrong to kill fetuses for that reason, for fetuses do not have experiences, activities, and projects to be continued or discontinued. Accordingly, the discontinuation account does not have the anti-abortion consequences that the value of a future-like-ours account has. Yet, it seems as intelligible as the value of a future-like-ours account, for when we think of what would be wrong with our being

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

killed, it does seem as if it is the discontinuation of what makes our lives worthwhile which makes killing us wrong.

Is the discontinuation account just as good an account as the value of a future-like-ours account? The discontinuation account will not be adequate at all, if it does not refer to the *value* of the experience that may be discontinued. One does not want the discontinuation account to make it wrong to kill a patient who begs for death and who is in severe pain that cannot be relieved short of killing. (I leave open the question of whether it is wrong for other reasons.) Accordingly, the discontinuation account must be more than a bare discontinuation account. It must make some reference to the positive value of the patient's experiences. But, by the same token, the value of a future-like-ours account cannot be a bare future account either. Just having a future surely does not itself rule out killing the above patient. This account must make some reference to the value of the patient's future experiences and projects also. Hence, both accounts involve the value of experiences, projects, and activities. So far we still have symmetry between the accounts.

The symmetry fades, however, when we focus on the time period of the value of the experiences, etc., which has moral consequences. Although both accounts leave open the possibility that the patient in our example may be killed, this possibility is left open only in virtue of the utterly bleak future for the patient. It makes no difference whether the patient's immediate past contains intolerable pain, or consists in being in a coma (which we can imagine is a situation of indifference), or consists in a life of value. If the patient's future is a future of value, we want our account to make it wrong to kill the patient. If the patient's future is intolerable, whatever his or her immediate past, we want our account to allow killing the patient. Obviously, then, it is the value of that patient's future which is doing the work in rendering the morality of killing the patient intelligible.

This being the case, it seems clear that whether one has immediate past experiences or not does no work in the explanation of what makes killing wrong. The addition the discontinuation account makes to the value of a human future account is otiose. Its addition to the value-of-a-future account plays no role at all in rendering intelligible the wrongness of killing. Therefore, it can be discarded with the discontinuation account of which it is a part.

IV.

The analysis of the previous section suggests that alternative general accounts of the wrongness of killing are either inadequate or unsuccessful in getting around the anti-abortion consequences of the value

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

of a future-like-ours argument. A different strategy for avoiding these anti-abortion consequences involves limiting the scope of the value of a future argument. More precisely, the strategy involves arguing that fetuses lack a property that is essential for the value-of-a-future argument (or for any anti-abortion argument) to apply to them.

One move of this sort is based upon the claim that a necessary condition of one's future being valuable is that one values it. Value implies a valuer. Given this one might argue that, since fetuses cannot value their futures, their futures are not valuable to them. Hence, it does not seriously wrong them deliberately to end their lives.

This move fails, however, because of some ambiguities. Let us assume that something cannot be of value unless it is valued by someone. This does not entail that my life is of no value unless it is valued by me. I may think, in a period of despair, that my future is of no worth whatsoever, but I may be wrong because others rightly see value—even great value—in it. Furthermore, my future can be valuable to me even if I do not value it. This is the case when a young person attempts suicide, but is rescued and goes on to significant human achievements. Such young people's futures are ultimately valuable to them, even though such futures do not seem to be valuable to them at the moment of attempted suicide. A fetus's future can be valuable to it in the same way. Accordingly, this attempt to limit the anti-abortion argument fails.

Another similar attempt to reject the anti-abortion position is based on Tooley's claim that an entity cannot possess the right to life unless it has the capacity to desire its continued existence. It follows that, since fetuses lack the conceptual capacity to desire to continue to live, they lack the right to life. Accordingly, Tooley concludes that abortion cannot be seriously prima facie wrong (*op. cit.*, pp. 46/7).

What could be the evidence for Tooley's basic claim? Tooley once argued that individuals have a prima facie right to what they desire and that the lack of the capacity to desire something undercuts the basis of one's right to it (*op. cit.*, pp. 44/5). This argument plainly will not succeed in the context of the analysis of this essay, however, since the point here is to establish the fetus's right to life on other grounds. Tooley's argument assumes that the right to life cannot be established in general on some basis other than the desire for life. This position was considered and rejected in the preceding section of this paper.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

One might attempt to defend Tooley's basic claim on the grounds that, because a fetus cannot apprehend continued life as a benefit, its continued life cannot be a benefit or cannot be something it has a right to or cannot be something that is in its interest. This might be defended in terms of the general proposition that, if an individual is literally incapable of caring about or taking an interest in some *X*, then one does not have a right to *X* or *X* is not a benefit or *X* is not something that is in one's interest.[11]

Each member of this family of claims seems to be open to objections. As John C. Stevens[12] has pointed out, one may have a right to be treated with a certain medical procedure (because of a health insurance policy one has purchased), even though one cannot conceive of the nature of the procedure. And, as Tooley himself has pointed out, persons who have been indoctrinated, or drugged, or rendered temporarily unconscious may be literally incapable of caring about or taking an interest in something that is in their interest or is something to which they have a right, or is something that benefits them. Hence, the Tooley claim that would restrict the scope of the value of a future-like-ours argument is undermined by counterexamples.[13]

Finally, Paul Bassen[14] has argued that, even though the prospects of an embryo might seem to be a basis for the wrongness of abortion, an embryo cannot be a victim and therefore cannot be wronged. An embryo cannot be a victim, he says, because it lacks sentience. His central argument for this seems to be that, even though plants and the permanently unconscious are alive, they clearly cannot be victims. What is the explanation of this? Bassen claims that the explanation is that their lives consist of mere metabolism and mere metabolism is not enough to ground victimizability. Mentation is required.

The problem with this attempt to establish the absence of victimizability is that both plants and the permanently unconscious clearly lack what Bassen calls "prospects" or what I have called "a future life like ours." Hence, it is surely open to one to argue that the real reason we believe plants and the permanently unconscious cannot be

[11] Donald VanDeVeer seems to think this is self-evident. See his "Whither Baby Doe?" in *Matters of Life and Death*, p. 233.

[12] "Must the Bearer of a Right Have the Concept of That to Which He Has a Right?" *Ethics*, XCV, 1 (1984):68–74.

[13] See Tooley again in "Abortion and Infanticide," pp. 47–49.

[14] "Present Sakes and Future Prospects: The Status of Early Abortion," *Philosophy and Public Affairs*, XI, 4 (1982):322–326.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

victims is that killing them cannot deprive them of a future life like ours; the real reason is not their absence of present mentation.

Bassen recognizes that his view is subject to this difficulty, and he recognizes that the case of children seems to support this difficulty, for "much of what we do for children is based on prospects." He argues, however, that, in the case of children and in other such cases, "potentiality comes into play only where victimizability has been secured on other grounds" (*ibid.*, p. 333).

Bassen's defense of his view is patently question-begging, since what is adequate to secure victimizability is exactly what is at issue. His examples do not support his own view against the thesis of this essay. Of course, embryos can be victims: when their lives are deliberately terminated, they are deprived of their futures of value, their prospects. This makes them victims, for it directly wrongs them.

The seeming plausibility of Bassen's view stems from the fact that paradigmatic cases of imagining someone as a victim involve empathy, and empathy requires mentation of the victim. The victims of flood, famine, rape, or child abuse are all persons with whom we can empathize. That empathy seems to be part of seeing them as victims.[15]

In spite of the strength of these examples, the attractive intuition that a situation in which there is victimization requires the possibility of empathy is subject to counterexamples. Consider a case that Bassen himself offers: "Posthumous obliteration of an author's work constitutes a misfortune for him only if he had wished his work to endure" (*op cit.*, p. 318). The conditions Bassen wishes to impose upon the possibility of being victimized here seem far too strong. Perhaps this author, due to his unrealistic standards of excellence and his low self-esteem, regarded his work as unworthy of survival, even though it possessed genuine literary merit. Destruction of such work would surely victimize its author. In such a case, empathy with the victim concerning the loss is clearly impossible.

Of course, Bassen does not make the possibility of empathy a necessary condition of victimizability; he requires only mentation. Hence, on Bassen's actual view, this author, as I have described him, can be a victim. The problem is that the basic intuition that renders Bassen's view plausible is missing in the author's case. In order to attempt to avoid counterexamples, Bassen has made his thesis too weak to be supported by the intuitions that suggested it.

[15] Note carefully the reasons he gives on the bottom of p. 316.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Even so, the mentation requirement on victimizability is still subject to counterexamples. Suppose a severe accident renders me totally unconscious for a month, after which I recover. Surely killing me while I am unconscious victimizes me, even though I am incapable of mentation during that time. It follows that Bassen's thesis fails. Apparently, attempts to restrict the value of a future-like-ours argument so that fetuses do not fall within its scope do not succeed.

V.

In this essay, it has been argued that the correct ethic of the wrongness of killing can be extended to fetal life and used to show that there is a strong presumption that any abortion is morally impermissible. If the ethic of killing adopted here entails, however, that contraception is also seriously immoral, then there would appear to be a difficulty with the analysis of this essay.

But this analysis does not entail that contraception is wrong. Of course, contraception prevents the actualization of a possible future of value. Hence, it follows from the claim that futures of value should be maximized that contraception is prima facie immoral. This obligation to maximize does not exist, however; furthermore, nothing in the ethics of killing in this paper entails that it does. The ethics of killing in this essay would entail that contraception is wrong only if something were denied a human future of value by contraception. Nothing at all is denied such a future by contraception, however.

Candidates for a subject of harm by contraception fall into four categories: (1) some sperm or other, (2) some ovum or other, (3) a sperm and an ovum separately, and (4) a sperm and an ovum together. Assigning the harm to some sperm is utterly arbitrary, for no reason can be given for making a sperm the subject of harm rather than an ovum. Assigning the harm to some ovum is utterly arbitrary, for no reason can be given for making an ovum the subject of harm rather than a sperm. One might attempt to avoid these problems by insisting that contraception deprives both the sperm and the ovum separately of a valuable future like ours. On this alternative, too many futures are lost. Contraception was supposed to be wrong, because it deprived us of one future of value, not two. One might attempt to avoid this problem by holding that contraception deprives the combination of sperm and ovum of a valuable future like ours. But here the definite article misleads. At the time of contraception, there are hundreds of millions of sperm, one (released) ovum and millions of possible combinations of all of these. There is no actual combination at all. Is the subject of the loss to be a merely possible

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

combination? Which one? This alternative does not yield an actual subject of harm either. Accordingly, the immorality of contraception is not entailed by the loss of a future-like-ours argument simply because there is no nonarbitrarily identifiable subject of the loss in the case of contraception.

VI.

The purpose of this essay has been to set out an argument for the serious presumptive wrongness of abortion subject to the assumption that the moral permissibility of abortion stands or falls on the moral status of the fetus. Since a fetus possesses a property, the possession of which in adult human beings is sufficient to make killing an adult human being wrong, abortion is wrong. This way of dealing with the problem of abortion seems superior to other approaches to the ethics of abortion, because it rests on an ethics of killing which is close to self-evident, because the crucial morally relevant property clearly applies to fetuses, and because the argument avoids the usual equivocations on 'human life', 'human being', or 'person'. The argument rests neither on religious claims nor on Papal dogma. It is not subject to the objection of "speciesism." Its soundness is compatible with the moral permissibility of euthanasia and contraception. It deals with our intuitions concerning young children.

Finally, this analysis can be viewed as resolving a standard problem—indeed, *the* standard problem—concerning the ethics of abortion. Clearly, it is wrong to kill adult human beings. Clearly, it is not wrong to end the life of some arbitrarily chosen single human cell. Fetuses seem to be like arbitrarily chosen human cells in some respects and like adult humans in other respects. The problem of the ethics of abortion is the problem of determining the fetal property that settles this moral controversy. The thesis of this essay is that the problem of the ethics of abortion, so understood, is solvable.

DON MARQUIS
University of Kansas

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

BOOK REVIEWS

Critique, Norm and Utopia. A Study of the Foundations of Critical Theory. SEYLA BENHABIB. New York: Columbia University Press, 1986. xiii, 455 p. Cloth $35.00, paper $15.00.

Seyla Benhabib's *Critique, Norm and Utopia* is a manifold achievement. The work presents a number of communitarian perspectives, eloquently demonstrating that communitarianism is not a univocal doctrine that simply asserts the primacy of the community over the individual. As an alternative to social contract theory, it forms a tradition in its own right in which the issue of (individual) autonomy is shown to be a persistent concern. The perspectives of the work can be characterized as historical reconstructions or narratives of (1) the idea of critique; (2) the background and development of Critical Theory; and (3) Jürgen Habermas's work as the culmination of (1) and (2).[1] Benhabib concurs with Habermas in criticizing "the philosophy of the subject" and the interrelated "work" model of action (10). Ultimately, she goes beyond Habermas, claiming that he does not adequately free himself from the philosophy of the subject in his own theory of communicative ethics.

Benhabib rejects the philosophy of the subject, because it entails several presuppositions, two of which are conspicuously Hegelian: (1) that history can be viewed as the work of a collective, singular subject that comes into being through externalizing itself in the historical process; and (2) that emancipation entails reappropriating the heritage of this subject (393). These aspects of the philosophy of the subject are manifest in the original programmatic statements of Critical Theory, and they are implicit in Habermas. It should be noted that Benhabib is not sympathetic to poststructuralist attacks on the philosophy of the subject. It is not the idea of subjectivity itself which needs to be abandoned, rather subjectivity conceived of as exclusively monological. Although Benhabib is prepared to defend "the dignity and autonomy of the rational subject" (15), she wants to argue that the rational subject is intersubjectively constituted and also that one should not presume that, by virtue of sharing rationality, all subjects have the same histories or interests.

She examines the idea of critique in order to establish the original

[1] Benhabib distinguishes her approach as "systematic" as opposed to straightforwardly historical. She also cautions against regarding Habermas's work in Hegelian terms as "last in line," and hence automatically "the most fully developed and the most adequate" (224).

0022-362X/89/8604/203–208 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

connection between critique and crisis, i.e., "a subjective evaluation or decision concerning a conflictual and controversial process" (19).[2] Hegel's rejection of the "mere criticism" of Kant and the Enlightenment aims to restore the aspect of crisis. This is necessary in Hegel's view, given the absence of unified ethical life in modern society. Benhabib describes two kinds of critique in the first two chapters, respectively, (1) immanent—"a non-criteriological inquiry, which allows its practitioners to criticize the opponents' arguments by showing their internal inconsistencies and contradictoriness" (9); and (2) defetishizing—"a procedure of showing that what appears as a given is in fact not a natural fact but a historically and socially formed reality" (21). Benhabib seems to regard these two kinds of critique as functioning in conjunction with each other, although not much is said about this. The main claim is that critique is crisis diagnosis, and that crisis diagnosis leads in two directions—to "norms" (the goal of "fulfillment") and to "utopia" (the goal of "transfiguration"). Following Habermas, she emphasizes a related distinction: that critique ideally includes an "explanatory-diagnostic" component as well as an "anticipatory–utopian" one. The former pertains to the perspective of the observer, while the latter pertains to the lived needs and experiences of social agents (142).

The subtitle of the work, "A Study of the Foundations of Critical Theory," indicates that Benhabib is concerned with tracing the emergence of Critical Theory, as formulated by the Frankfurt School, out of its background in Hegel and Marx. The "transformation" of the idea of critique, the title of part II, amends the Marxist idea of critique of political economy to the "critique of instrumental reason." The latter has the advantage of not excluding consideration of cultural and psychological relations that individuals actually experience. Benhabib goes on to claim that the development of Critical Theory resulted in the abandonment of the explanatory-diagnostic component of critique. This occurred as it allied itself more and more with aesthetics as the (utopian) alternative to instrumental reason, beginning with *Dialectic of Enlightenment*.[3] While Critical Theory attempted to break free from the model of action inherited from Hegel and Marx, it still held onto the presuppositions of the philosophy of the subject (213).[4]

[2] The discussion of critique and crisis stems from Reinhart Kosseleck's *Kritik und Krise* (Frankfurt: Suhrkamp, 1976). Also see: Habermas's *Legitimation Crisis*, Thomas A. McCarthy, trans. (Boston: Beacon, 1975), pt. I.

[3] Theodor Adorno and Max Horkheimer (New York: Seabury, 1972).

[4] Benhabib argues that the model of action in Hegel and Marx is the work model in which a subject externalizes and produces itself through labor. It also shows the

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Habermas rejuvenates Critical Theory by positing communicative reason as an alternative to instrumental reason. Rather than merely reject the work model of action, he replaces it by the model of communicative interaction which is guided by ideal speech conditions. Critical Theory's original aim of social transformation is insured, for both components of critique are included: the explanatory-diagnostic (which reestablishes a bridge between philosophy and the social sciences) and the anticipatory-utopian (which distinguishes Critical Theory from functionalist theories, since it preserves an interest in the emancipation of those who actually suffer).

Benhabib makes no secret that her projects reiterate stories Habermas has told. Part I on the "origins" of critique, which traces Hegel's critique of Kant's critique(s) and Marx's critique of Hegel, reads like the first part of Habermas's *Knowledge and Human Interests,*[5] rewritten with his later self-criticisms in mind.[6] I want to dwell upon two examples that reveal that Benhabib relies rather too heavily upon Habermas.

The first example concerns the interpretation of Hegel's development. Benhabib follows Habermas in beginning the story with *Natural Law* (written before 1802), which both she and Habermas see as being burdened by the model of the polis and the example of Greek tragedy. Hegel does not formulate a realistic alternative to address the modern dilemma of the individual's relation to society. The story moves on to the *Jena Realphilosophie* (1805/6), which highlights labor and interaction, and promisingly turns away from the philosophy of the monological subject. In the *Phenomenology of Spirit* (1807), the promise of transcending monological subjectivity founders, owing to the priority given to transsubjectivity, the vantage point of Spirit. Benhabib's Habermasian version fails to account for

---

influence of expressivism, a model which she borrows from Charles Taylor. Unfortunately, Benhabib does not explore contradictions between the work and expressivist models. Taylor emphasizes that the latter is an aesthetic concept (it includes an element of play) and contains demands for unity, freedom, and communion with man and nature. See Taylor, *Hegel* (New York: Cambridge, 1975), pp. 24–29; 38.

[5] Jeremy J. Shapiro, trans. (Boston: Beacon, 1971).

[6] For Habermas's self-criticisms, see (1) "A Postscript to *Knowledge and Human Interests,*" in *Philosophy of the Social Sciences,* III (1973): 157–189; and (2) "A Reply to my Critics," in *Habermas: Critical Debates,* John Thompson and David Held, eds. (Cambridge: MIT, 1982), pp. 219–283. For his criticism of the philosophy of the subject, see (1) *The Theory of Communicative Action,* vol. I: *Reason and the Rationalization of Society,* Thomas A. McCarthy, trans. (Boston: Beacon, 1984), pp. 387f.; and (2) *The Philosophical Discourse of Modernity: Twelve Lectures,* Frederick Lawrence, trans. (Cambridge: MIT, 1987), pp. 294–326. Also see Richard Bernstein's useful introduction in *Habermas and Modernity,* Bernstein, ed. (Cambridge: MIT, 1985).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

several matters: (1) that the emphasis upon Greek tragedy in *Natural Law* can itself be regarded as an interest in intersubjectivity; (2) that Hegel certainly dwells upon intersubjectivity already in *System of Ethical Life* (1802/3) and also in *First Philosophy of Spirit* (1803/4); and (3) that it is possible to argue that the *Phenomenology* retains an interest in intersubjectivity, following the example of Greek tragedy, through the relation of the reader/spectator reacting to the "performance" of the text/drama.

The second example of Benhabib's over reliance upon Habermas concerns the interpretation of Horkheimer and Adorno's concept of mimesis. Benhabib follows Habermas in criticizing it as an aesthetic concept. In turning the work model upside down, mimesis is a kind of narcissistic mirror which ignores the crucial role that linguistic communication plays and which fails to formulate a viable alternative model. Benhabib is misled by Habermas's assessment of mimesis as part of the overall pessimism of Horkheimer and Adorno (both acknowledge that their pessimism was affected by the historical moment in 1944). Both Benhabib and Habermas ignore that mimesis can be construed as providing a positive model of human interaction. It provides a contrast to "false projection," where an individual fails to be able to acknowledge the other as other and regards the other as like herself. False projection is a pathological form of relation which Horkheimer and Adorno invoke in order to understand anti-Semitism and fascism (*op. cit.*, p. 187).

Mimesis includes Aristotelian connotations of being: an imitation of action, a natural extension from childhood pleasure, and potentially a learning experience (*Poetics*, 1448a-b). It also includes a psychoanalytic premise: that one must move beyond the child's intense identification with the parents to a new kind of bond, which, in not obscuring or denying difference, fosters autonomy. Mimesis provides a healthy alternative—at worst, a glimmer of hope in the darkness and, at best, an astute concrete analysis of a social phenomenon that was of crisis proportions.

Benhabib's discussion of Habermas is complex and enormously informative. I shall not be able to do it justice here. She is skeptical of Habermas's Kantian tendencies, discernable in his attention to transcendental argument (267), proceduralism (288), and universalist egalitarian commitments (306). She also worries that Habermas does not successfully carry forth an anticipatory-utopian component. In allying himself with "a democratic public ethos," he is opting for the goal of fulfillment over transformation (254;283). She lauds Habermas for his interest in rights and entitlements, while she criticizes him for neglecting issues of needs and solidarity (339).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

At the heart of Benhabib's criticism of Habermas is the charge that he ignores natural, social, and cultural differences. She rejects the possibility of a generalized other, which Habermas takes from G. H. Mead and which bears the mark of the philosophy of the subject. Benhabib offers instead the notion of a concrete other, which she takes from Carol Gilligan. The concrete other insists upon "each and every rational being as an individual with a concrete history, identity, and affective-emotional constitution" (341). The concrete other is identified with a feminist perspective, although its relevance is not limited to gender issues (409). Benhabib's notion of the concrete other bases itself upon a norm of "complementary reciprocity." She emphasizes that the context for human interaction is often private and noninstitutional, and thus there is the need for norms of solidarity, friendship, love, and care.[7]

One might well link Benhabib's notion of the concrete other to Horkheimer and Adorno's concept of mimesis, since they portray the Jews as concrete others. [Horkheimer and Adorno also mention a connection in their analysis between Jews and women (*op. cit.*, p. 112; see also p. 248).] It is not easy to reconcile the notion of the concrete other with Habermas's views. For example, Benhabib does not deal with the way in which love and care are not bound by linguistic communication. It is also unclear why Benhabib would want to endorse Habermas's exclusivist bridging of philosophy and the social sciences, since it is predicated upon the Kantian separation of the theoretical, the practical, and the aesthetic (and the implicit devaluing of the latter). Is not aesthetic discourse precisely the arena where we find the concrete other represented and taken seriously? Certainly aesthetic discourse is likely to be more articulate about and attentive to our "affective-emotional" selves.

Benhabib proposes that Critical Theory constitutes a third alternative in contemporary philosophy, situated between neo-Kantianism and contextualism. Like neo-Kantians, she is not prepared to do away with standards of validity (or the model of the rational subject). Like the contextualists, she accepts "the inevitable cultural, historical hermeneutical and ontological presuppositions which form the horizon of our standpoint" (15). In reaction to the neo-Kantian

[7] For a view that is more bluntly critical of Habermas's failure to take account of such norms, see Nancy Fraser, "What's Critical about Critical Theory? The Case of Habermas and Gender," in *Feminism as Critique*, Benhabib and Drucilla Cornell, eds. (Minneapolis: Minnesota UP, 1987). Benhabib's article in the same volume, "The Generalized and the Concrete Other: The Kohlberg-Gilligan Controversy and Feminist Theory," offers important clarifications concerning the notion of the concrete other.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

elements of Habermas's version of Critical Theory, Benhabib introduces her own version, which seems to lean more toward contextualism. Perhaps we might consider this as the basis of a fourth, rather than an extension of a third, alternative.

ELLIOT L. JURIST

University of California/Santa Barbara

Ought, Reasons, and Morality. W. D. FALK. Ithaca: Cornell University Press, 1986. 291 p. $29.95.

W. D. Falk is a hedgehog, to use Isaiah Berlin's term. Over the past forty years he has written a series of papers which grapple with the most fundamental problems of moral philosophy. What is value, and how is it known? How is moral obligation possible? How, that is, can rational agents be bound by morality in a way which is inescapable and which gives them rationally conclusive motives? How can reason be practical? How can conduct be *guided* by reason, not just in the sense that it proceeds in the knowledge of reasons, but shaped by an awareness that they are reasons? And how is that awareness itself possible? How can practical thinking go validly from *is*s to *ought*s, even though no practical judgment follows deductively from the premises on which it is based?

Given the breathtaking range of these questions, it may seem odd to characterize Falk as a hedgehog. Only the nimble fox could hope to engage them all, and, even then, only in his relatively superficial, foxlike way. It is not too much of an exaggeration to say, however, that, for Falk, all these questions are but different aspects of the same underlying problematic, different formulations of a single-root issue. Whether he is concerned with the nature of value and its epistemology, or with the relation between moral obligation and the autonomy of the moral subject, Falk takes himself to be exercised with a question that is, at bottom, much the same, even though it may appear in different dress, with different emphases and stress. And it has been almost always to expose, engage, and, finally, to solve this single-root issue of moral philosophy that Falk has been burrowing.

The present book is a very welcome collection of Falk's essays. Most have been previously published, and many of these with great impact, but five of the essays are new. The collection is welcome, of course, partly because of this new material; "Fact, Value, and Nonnatural Predication" is an important piece, for example. But the

0022-362X/89/8604/208–214 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

collection is welcome also because almost all the valuable collections of midcentury moral philosophy, in which Falk's work was frequently represented, are now out of print. In recent years, only the determined scholar or student was likely to come upon such important pieces as " 'Ought' and Motivation" or "Morality, Self, and Others." Finally, by bringing Falk's work together in one place, this book shows what might otherwise not have been so apparent, the hedgehog burrowing, often from different starting places, but almost always to the same root.

Many of Falk's philosophical preoccupations, as well as the main lines of his own positive views, were apparent in the earliest piece here collected, "Morals without Faith," first published in 1944. As the title indicates, his problem there is whether and how there can be moral obligation without God. Those who know Falk's later work on moral obligation, for example, "Morality, Self, and Others," published almost twenty years afterward, will be amazed at how much of this later work is here foreshadowed. God cannot provide the requisite support for moral obligation, since the latter requires an *internally* given 'ought', a demand which, although experienced as external, because inescapable, nonetheless has its source within the self. But how is this possible? Its possibility, Falk believes, derives from a fundamental distinction between what a person happens to want at a given moment, and what that person would want were he to engage in a kind of thinking which can be undertaken only by a creature with powers of reflection and judgment: the careful and judicious mulling over of alternatives for all that he can know them to be. The idea that an alternative is one that "would attract us if we formed, as Spinoza said, an *adequate idea* of it," (172) is enough to fund the notion that the alternative is worthy of choice, that it *ought* to be chosen, other things equal. But this 'ought', Falk thinks, is not yet the moral 'ought'. The latter has a "categorical practical necessity"; it says what one must do in preference to any other alternative. But here again we can distinguish between simply happening to prefer one alternative over others, and its being the case that a person would prefer a given alternative over all others were he to give them all an adequately judicious review. This provides all the resources needed to fund a moral 'ought', since it gives a basis for the idea that an alternative is "unalterably forced upon us by the thought of its own nature and effects, and despite any sacrifice involved in doing it" (174). The special practical necessity of the moral 'ought' is thus tied to the availability of a conclusive motive, one that would actually weigh most heavily with a person were she adequately to exercise the uniquely rational capacity for reflection and judg-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ment. And why, it may be asked, should we care about exercising this capacity other than, perhaps, out of a kind of solidarity with the rational species? Here Falk anticipates the suggestive line of "Morality, Self, and Others" that such "rational self-determination" is the condition of continuing self–identity as "the unified, distinct, and enduring subject of our own feeling-states."

> Self-identity or unity of consciousness is not a self-subsisting thing: it is constantly in jeopardy, increased or diminished by the extent to which we offer resistance to our impulses and view them before we follow them in relation to the true nature of their objects and to one another (177).

All of this already in 1944.

I shall want to say some things about this general approach, but before I do, I want, first, to show how it resurfaces again and again throughout the essays. This is most obvious in the other pieces that follow "Morals without Faith" in Part Two of the book. All these pieces concern the nature of morality in some way or other, and it is a recurrent theme in them that, although part of our thinking about morality is that it concerns how we should act in relation to others, we are also inclined to think that morality consists of "all valid directives for action; about all those things which a person might not otherwise do readily but which, for good and compelling reasons in the nature of the case, he ought to do and would have to break himself into doing whether for the sake of others or his own" (204). Moreover, Falk repeatedly insists, this is what must be fundamental. Only this idea can ground an 'ought'. It is but contingently true, albeit deeply so, that caring for others, or even refraining from doing them harm, is among the things we ought to do.

The same theme arises in "On Learning About Reasons," new in this volume, and the well–known "Action-Guiding Reasons." Here Falk attempts to occupy a middle position between those who hold a purely instrumental conception of rationality and those who maintain the possibility of direct cognitive access to truths about rational ends. Against the former, Falk argues that practical thinking involves careful review of what can be known of alternative ends as well as assimilating information about alternative means to ends. Again, we can distinguish between simply happening to have an end and an end's being the result of having what one knows about it "comprehendingly before one." In the second case, we can say that the end is supported by reason, since it is supported by the kind of thinking of which only a rational creature is capable. Against the intuitionists, Falk argues that what settles the truth about what reasons there are

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

for a person to act can be nothing independent of what considerations he would himself come to be moved by through knowledgeable, experienced, and judicious review in the limit.

Here we have the same view, put now in the language of reasons rather than the language of morals. For Falk, there are not two species of 'ought', one moral and one rational, but only one. What a person ought to do, all things considered, depends on what there are the best reasons for him to do. And what there are the best reasons for him to do depends on which alternative he would come to prefer through experienced and knowledgeable review in the limit.

Falk turns, in two essays, to attempting to exhibit the essentials of this view in Hume, both as an account of practical reasoning and as an account of the relation between 'is' and 'ought'—in short, to arguing that Hume was digging in the same burrow, and to much the same effect. Falk's exegetical case is interesting, and a useful corrective to crude instrumentalist readings of Hume on reason, and to simplistic naturalist or noncognitivist interpretations of Hume on the moral judgment.[1] But one does have the feeling that the text is being selected to present a Falkian Hume. My concern is not so much that Falk rests a great deal on passages from the *Enquiry* that take a more extensive view of the role of reason, and on "a reading of the *Treatise* in the light of it." It is rather that what Hume is explicitly discussing in these passages is the role of reason in the making of judgments of personal merit, of virtue and vice, and not of alternatives directly before an agent. And it is not clear what direct conclusions can be drawn from these passages about reasons to *act* and 'ought to *do*s. In fact, I think much can be filled in which would be helpful to Falk's line. So, for example, Hume,[2] largely following Hutcheson, holds the following views: first, that there is a species of obligation, moral obligation, consisting in the approbation and disapprobation of moral sentiment; second, that, although such sentiment concerns character and motive in the first instance, it is also true; and, third, that there is a (moral) obligation or duty to do what the virtuous person would do and the vicious person would not. But, at the very least, there are more argumentative steps needed to make Falk's case, and, I think, questions still remain about whether many of the passages Falk quotes can be made to do the general work he wants of them.

[1] I am not saying either that any account of Hume's view of reason as instrumental must be crude, or that any account of Hume's metaethics as naturalist or emotivist must be simplistic.

[2] *A Treatise of Human Nature*, L. A. Selby-Bigge, ed., with text revised and notes by P. H. Nidditch (New York: Oxford, 1978), pp. 478/9.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Of the papers that are new in this volume, the most substantial is "Fact, Value, and Nonnatural Prediction." Falk's point of departure here is G. E. Moore's famous, but mysterious and rarely understood, doctrine that value is a nonnatural property. In a penetrating treatment of Moore's discussion of this idea in *Principia*, "The Conception of Intrinsic Value," and "Reply to My Critics" from the Schilpp volume, Falk demonstrates that the substance of Moore's contrast was not between properties that are empirically discoverable and those which are not, but between properties that are intrinsic to the valuable object and properties that, while depending only on the object's intrinsic properties, are not themselves intrinsic to the object. Moore's position, then, was that the sort of value of fundamental importance for ethics, intrinsic value, depends entirely on the intrinsic nature of what has it, but is not itself part of its intrinsic nature.

Falk takes great pains in tracking Moore's forbidding dialectic, because he thinks Moore saw deeply into the nature of value and almost got it right. Moore was right, he argues persuasively, to think that the sort of value of fundamental importance for ethics is not part of the intrinsic nature of valuable things. And Moore was almost right in thinking that such values depend entirely on the intrinsic nature of what has them. Where Moore went wrong was in misconceiving the way in which an object's value depends on its intrinsic nature. Something has value *on account of* its intrinsic nature, where 'acount' should be taken literally. The dependence of value on a thing's intrinsic nature is not direct. It consists, and is not simply revealed, in the fact that something would evoke and sustain favor in an appropriately experienced, sensitive, and diligent reviewer through an adequately comprehending accounting of the thing's intrinsic nature. Indeed, this explains why it is that value cannot itself be an intrinsic property. The appropriately comprehending favor in the limit which is criterial of value is itself comprehensive of a thing's whole intrinsic nature.

The part that Moore got wrong is, of course, the part supplied by the general Falkian view and, more specifically, by Falk's Hume. And, it must be said, if one takes the *Enquiry* as advancing a view about value generally, and not just about personal merit, the similarities are startling.[3] Thus, regarding physical (rather than moral) beauty: "the beauty is not in any of the parts or members of a pillar but results from the whole when that complicated figure is presented

[3] With the major dissimilarity that, for Hume, the judgment of personal merit is not of a trait's intrinsic nature, but almost entirely of its usual effects.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

to an intelligent mind susceptible to those finer sensations."[4] And: "the crime or immorality is no particular fact or relation which can be the object of the understanding, but arises entirely from the sentiment of disapprobation which, by the structure of human nature, we unavoidably feel on the apprehension of barbarity or treachery"[5]—and, not by any simple apprehension, but only one adequately prepared by the full use of reason in the limit.

Even if one agrees with Falk, it is probably misleading to characterize Moore as "fail[ing] to grasp" the kind of dependence value has on fact "though it was before him all along." Moore considered such views, and rejected them. In fact, part of his motivation in writing "The Conception of Intrinsic Value" was to characterize a sort of value which does not depend on whether "some particular individual or class of individuals either actually has, *or would, under certain circumstances,* have a certain kind of mental attitude towards the thing in question."[6] Whether any such value exists is, of course, another matter.

Falk's burrowing has led him over forty years to return consistently to the same position. But what about the position itself? Plainly it is a very deep and powerful view. It gives Falk a unified perspective on the most fundamental problems of moral philosophy. But is it true?

I am disposed to accept much of it. It must be right, it seems to me, that there could be no truths about what reasons there are for a person to act which are independent of what the person would find motivating were he to consider things rationally. What I am unsure of is whether rational consideration and deliberation can itself be fully understood in terms of a continued mulling over of what can be known in the limit. One way into what worries me is to notice that there may simply not be any stable preference in the limit in many cases, partly because consideration can be from different points of view that are not apt to be stably integrated simply through continued consideration, even in the limit. Take, for example, the much–discussed contrast between the point of view of one person among others and a person's own personal standpoint—a contrast which is undoubtedly much too crude (in both directions). Is it obvious that deliberation that takes account of both of these points of view will

[4] *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, L. A. Selby-Bigge, ed., with text revised and noted by P. H. Nidditch (New York: Oxford, 1985), p. 292. See also p. 290.
[5] *Ibid.*
[6] "The Conception of Intrinsic Value," in *Philosophical Studies* (London: Routledge & Kegan Paul, 1948), p. 254.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

stabilize, even on a given occasion, in the sense that continued consideration would not alter the preferred line of conduct? On the contrary, might it not happen that continued consideration of both points of view leaves no stable preference, with choice depending, perhaps, on such matters as which point of view was considered last?

On the other hand, if preference does stabilize, this may simply be because one point of view, the agent's own, is motivationally dominant. Although other points of view are considered, and come to inform what the agent wants from his own standpoint, nonetheless, because the "all things considered" standpoint is the agent's standpoint, albeit as fully informed, what he judges best will then depend on what he values most from his own (perhaps enlarged) point of view. Since Falk rejects an instrumental conception that would simply identify rational action with what best serves the agent's ends, and maintains that an agent's ends can themselves be rationally criticized, the question arises: Why believe that the final standpoint of rational consideration must be the agent's own (informed) standpoint? After all, this is by no means the only standpoint from which reflective motivation is possible. Moreover, there is one good reason for not believing this. An autonomous rational being, disposed to review critically even his conception of rationality, will realize that in considering what he has reason to do on a given occasion he is ultimately raising the question of what it would be rational for a person such as him to do. And the value things have for *him* (as opposed to for *the agent*) is irrelevant to this question.

STEPHEN L. DARWALL

University of Michigan/Ann Arbor

Kant and the Claims of Knowledge. PAUL GUYER. New York: Cambridge University Press, 1987. 482 p. Cloth $59.50, paper $19.95.

Although a lengthy book, extraordinarily rich in detail, its plot can be summarized briefly. The story begins in the mid 1770s with the *Duisburg Nachlass,* which contains Kant's earliest known efforts to work out his theory of experience; it ends circa 1790 with the latest of the *Reflexionen* devoted to the emendation of the "Refutation of Idealism." The major claim is that, with the exception of the "Analogies" and the "Refutation," virtually all that lies between constitute

0022-362X/89/8604/214–221 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

failed attempts to achieve a "transcendental theory of experience." The successful theory, as it finally emerges, reduces to the claim that knowledge of even the subjective succession of one's representations requires knowledge of a law-governed realm of objects distinct from the self. Moreover, this requirement is said to entail a set of restrictions, specifying the conditions that must be satisfied if experience, of even the subjective sort, is to be possible, rather than a set of absolutely necessary conditions imposed by the mind on appearances. Thus, both transcendental idealism and the Transcendental Deduction are dismissed. What remains is merely this set of conditionally necessary constraints on experience, which presumably suffices to answer a Cartesian skepticism about the external world.

What makes this book distinctive are the single–mindedness and erudition with which Guyer pursues his thesis. He leaves no *Vorlesung* unmined, no *Reflexion* unmentioned, that can possibly bear on the topic. In fact, it almost seems as if Guyer's Kant kept his best thoughts for himself and his students and published only his mistakes and confusions. Unfortunately, for all his prodigious scholarship, Guyer's central thesis is neither compelling nor novel. Reduced to its essentials, it amounts to a restatement of the familiar Strawsonian account of "what is living and what is dead" in Kant. In what follows, I shall summarize and comment briefly on the contents of each of the five parts of the work.

The first part deals mainly with the *Duisburg Nachlass*. Guyer here makes two major claims: (1) The texts contain two different strategies for establishing a priori principles (which include only those which later became the analogies): one from the conditions of objective judgments, the other from the nature of apperception. (2) Kant is ambivalent regarding both the kind of necessity assigned to these principles and the function they supposedly serve. Usually it is a merely conditional necessity (if we are to have experience, then these principles must hold) and their function is to restrict the kinds of experience possible for us. Sometimes it is an absolute, unconditioned necessity; and, on this view, their function is to constitute experience.

The obscurity of the texts make it difficult to evaluate Guyer's claims. Nevertheless, on several points his interpretations seem questionable. The most important example is his discussion of a passage in R 4678, in which Kant characterizes the principles as "restrictions" [*Restrictionen*] rather than "principles" [*Grundsätze*]. Predictably, Guyer takes this to be tantamount to the distinction between conditional and unconditional necessity (54/5); but the text

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

does not indicate that Kant had any such contrast in mind. On the contrary, the most natural reading is that the validity of the rules in question is limited to objects of possible experience as opposed to objects in general.[1] Clearly, it is one thing to claim that the scope of these rules is restricted to objects of possible experience and quite another to contend that they apply only to the extent to which the experience of independently existing objects is in fact possible (with this extent being an open, contingent matter). Pace Guyer, I see no evidence supporting the thesis that the latter is Kant's view.

Guyer's analysis of the Transcendental Deduction, the subject of part two, builds on the account of Kant's uncertainty regarding his basic premises. Because of this uncertainty, he claims that "it was inevitable that the deduction . . . never amounted to more than a disjointed summary of significantly different strategies" (73). As before, the basic division is between strategies that assume a strong premise about knowledge of objects and those which start with an assumption about apperception or self-knowledge. He complicates matters a bit, however, by introducing a further distinction between strategies that presuppose some a priori knowledge and those which do not. This yields the following possibilities: (IA) judgments about empirical objects are possible, and contain some synthetic a priori knowledge, which implies further a priori knowledge of the categories; (IB) although such judgments do not themselves assert any claims to a priori knowledge, they nevertheless presuppose a priori knowledge of the categories; (IIA) the possibility of apperception, as itself a kind of synthetic a priori knowledge, implies a priori knowledge of the categories; (IIB) the possibility of apperception, even as a form of merely empirical self-knowledge, implies a priori knowledge of the application of the categories to objects existing independently of the self (85/6).

The essential feature of this account is Guyer's view of apperception and its function in Kant's unsuccessful arguments. Following Dieter Henrich, he takes Kant's basic premise to be that the self has a priori knowledge (with "Cartesian certainty") of its numerical identity over time (IIA), from which it is then inferred that such knowledge requires a synthesis according to the categories. Unlike Henrich, however, he contends that this argument suffers from a

[1] This concurs with the interpretation offered by Theodor Haering, *Der Duisburg'sche Nachlass und Kants Kritizismus um 1775* (Tübingen, 1910), p 51, which is the standard monograph on these texts. Although Guyer refers to this work in his notes, he gives no evidence of having consulted it.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

fateful ambiguity in the basic premise. It can be taken either as the unobjectionable analytic claim, expressing a conditional necessity, that, if I am to become conscious of my identity through time, then my representations must be synthesized, or as the problematic synthetic claim, expressing an unconditional necessity, that it must always be possible for me to become conscious of my identity. Only the latter reading licenses the inference to the necessity for a special a priori synthesis, which is in turn the basis for the necessity of the categories. But, since there is no reason to accept this premise, the argument for the categories collapses. Moreover, since the proof-strategies assuming knowledge of objects employ a premise that the skeptic need not accept, only IIB, which is worked out only in the "Refutation of Idealism," remains viable.

Much can be said in response to Guyer's demolition of the deduction, but two points must here suffice. First, like all patchwork theorists, Guyer frequently goes out of his way to "deconstruct" arguments into separate strands only superficially connected, when it seems perfectly reasonable to regard these strands as organically related stages in a single argument. The clearest example of this is his account of the apperception and the objectivity components of the B–Deduction (115). Second, as I have argued elsewhere, the evidence for his interpretation of Kant's apperception doctrine as affirming a privileged kind of self-knowledge is both slight and ambiguous.[2] Moreover, once this reading is rejected, Guyer's critique loses all its force; since we are no longer required to take Kant to be affirming either an unconditional necessity regarding the consciousness of one's identity through time or the necessity of a distinct transcendental synthesis (as opposed to an empirical synthesis governed by a priori rules, which Guyer accepts).

Parts three and four, which deal with the "Analytic of Principles" and "Refutation of Idealism," respectively, are the most interesting portions of the book. Guyer has worthwhile things to say about the "Schematism" and the "Mathematical Principles," but the bulk of his attention is appropriately devoted to the "Analogies." The strength of his interpretation is the account of their interdepen-

[2] See "Reflections on the B-Deduction," Spindel Conference 1986: The B-Deduction, *Southern Journal of Philosophy*, XXV, supplement, pp. 12–14. Although I was there arguing against the earlier version of Guyer's analysis of the deduction ["Kant's Tactics in the Transcendental Deduction," *Essays on Kant's Critique of Pure Reason*, J. N. Mohanty and R. W. Shaker, eds. (Norman: Oklahoma UP, 1982), pp. 157–199], identical considerations apply to the book, since his central thesis remains the same.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

dence. He argues convincingly that all three are required for any empirical time determination whatsoever. He also provides valuable discussions of permanence, the unperceivability of time, and the simultaneity of cause and effect, as well as telling criticisms of many influential interpretations. But, in spite of these virtues, his account is not without problems of its own.

Most of these concern the Second Analogy, where the distinctive feature of Guyer's analysis is his insistence that even the merely subjective order of representations is "interpreted" rather than immediately apprehended. This becomes a crucial hidden premise on Guyer's reading, because it is necessary to appeal to it in order to disarm the objection that reference to subjective succession suffices to determine the occurrence of an objective change (254/5). The key point is that the determination of the true subjective order, say that rep *B* is a present perception of *B* and rep *A* a memory of a previous experience of *A*, can be decided only on the basis of knowledge of the objective succession *A–B*; but, since this sequence is likewise not directly accessible, "it can be inferred only from a law which dictates that, in given circumstances, B must have succeeded A" (257).

As Guyer recognizes, this appears to conflict with the requirement that the causal laws to which we supposedly appeal in order to determine the temporal order of our experience must themselves have been derived inductively. How can empirical laws serve as necessary conditions of experience? Guyer's response turns on a deflation of the notion of being a condition of possible experience to that of being a condition of the confirmation of particular judgments regarding temporal order. Thus, eschewing any metaphysical reading, according to which the analogies function in the constitution of an objective temporal order, he insists that they are concerned exclusively with determining the epistemological framework for the confirmation of such judgments. Within this purely epistemological framework, he argues, there is no incoherence or circularity in the notion that causal laws are required to justify claims about objective and even subjective succession, even though it is only by means of such succession that we acquire knowledge of causal laws. A problem would arise only if the very same sequence is both derived from a particular causal law and employed as evidence for that law (258/9, 315, 326).

But, if not incoherent, the proposition that I can confirm a judgment of temporal succession only if I can subsume this succession under a known causal law is plausible neither in its own right nor as an interpretation of Kant. Does it really make sense to claim that "it

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

is only if we are in possession of causal laws which dictate that in the relevant circumstances—that is, not in general, but in the particular circumstances of wind, tide, setting of the sails, and so forth, which are assumed to obtain—the ship could only sail downstream that we actually have sufficient evidence to interpret our representations of it to mean that it is sailing downstream" (252)? Leaving aside the fact that the interpretation of our representations would seem to be quite a different matter from the confirmation of judgment claims, this makes the requirements for the determination of temporal succession quite strict, indeed. Correlatively, should the possibility of arriving at causal laws inductively be taken as a premise of the Second Analogy, rather than as a result that the argument is intended to help ground?

Given his central thesis, Guyer's account of the Refutation of Idealism is the focal point of his book. Unlike the original version, he insists that the 1787 Refutation involves no idealistic assumptions or implications. The successful argument is not found in the *Critique,* however, but only in the subsequent *Reflexionen.* In essence, this argument involves showing that subjective succession can be made determinate only if it is regarded as correlated with and caused by objects conceived to exist independently of the self and its states. It is this argument (strategy IIB) which supposedly represents "nothing less than Kant's final attempt to solve the puzzle about the objective validity of the a priori concepts of understanding first discovered in 1772" (297).

Since it is impossible to do justice here to Guyer's intricate analysis, three brief comments must suffice. First, in spite of his analysis, it remains implausible to take the Refutation of Idealism as playing a central role in establishing the categories. Second, the allegedly successful argument is found only in a portion of a single *Reflexion* (6313), which also contains other, quite different, lines of thought. Finally, since Guyer's reading effectively makes the Refutation of Idealism into a special application of the Second Analogy, while also supposedly justifying an essential premise of that argument, the two must stand or fall together. Consequently, one's view of the present argument will probably depend on one's evaluation of Guyer's success in avoiding the difficulties raised by his interpretation of the Second Analogy.

Having shown the irrelevance of idealism to Kant's positive achievement in the first four parts, Guyer turns in the fifth to a critique of this idealism. Rejecting all efforts (including my own) to depict transcendental idealism as "an anodyne conceptual analysis,"

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

he characterizes it as the dogmatic claim that things in themselves are nonspatial or temporal. He also distinguishes three attempts in the *Critique* to argue for this dogma, all of which are unsuccessful. These are the direct arguments in the "Aesthetic" (from the nature of space and time) and in the "Analytic" (from the nature of judgment), and the indirect argument in the "Antinomy." I shall here limit myself to a consideration of the first.

Whereas most interpreters take Kant to move from the premise that space and time are forms of sensibility to the conclusion that things in themselves are not spatial or temporal, Guyer insists that the argument proceeds in precisely the opposite direction. Since he contends both that Kant does in fact and must argue in this way, his two theses may be considered separately. Unfortunately, the evidence for his first claim rests largely on the order of the two conclusions drawn from the Metaphysical and Transcendental Expositions of space. Instead of taking these as complementary conclusions, mutually entailed by the arguments, which is certainly the most natural reading, he thinks that Kant intended only the first conclusion—"Space does not represent any property of things in themselves"—to follow directly from the Expositions and the second—"Space is nothing but the form of all appearances of outer sense"—to follow from the first. Guyer does cite a number of other passages from the *Nachlass* in support of this reading; but all that they do is reiterate the familiar Kantian thesis that a priori knowledge of space is incompatible with the assumption that it is derived from things in themselves. As Guyer himself recognizes (358), this does not of itself entail that things in themselves are not spatial.

Guyer's main claim, however, is that Kant must argue in this way, because his only available premise is the incompatibility of a priori knowledge of a determination with its independent existence (362). Moreover, he insists that this premise is itself the result of a failure to distinguish between conditional and absolute necessity. On his view, all that may be inferred from our supposed a priori knowledge of space is the conditional claim: "Necessarily, if we are to perceive an object then it is spatial and euclidean." But instead of drawing this modest conclusion, Kant's mistaken view of a priori knowledge led him to assert the absolute necessity: "If we are to perceive an object, it is necessarily spatial and euclidean." Since this, unlike the former, is incompatible with the supposition that things in themselves are spatial, it leads directly to idealism.

Although it is crucial for Guyer to show that a priori knowledge

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

requires only a conditional necessity, his supporting argument reduces to the bare assertion that "it seems at least possible to imagine that we could know because of certain constraints on our ability to perceive that any object we perceive must have a certain property" (363). Kant had good reason to reject this kind of necessity, however, since it is not only conditional but subjective. Guyer acknowledges this, but he still insists that Kant's only basis for rejecting a merely subjective necessity is his unwarranted assumption that a priori knowledge requires the absolute variety. But this totally misconstrues Kant's position. He does not reject such a necessity because it entails that a priori knowledge applies only contingently to objects (Guyer's view), but because it entails that it does not apply at all. At best, it yields a knowledge of the "natural constitution" of our cognitive faculties, not a knowledge of the world, either as it is in itself or as it appears.

Finally, it does not follow from the fact that Kant rejects the kind of necessity advocated by Guyer, that he accepts an absolute, unconditioned necessity. On the contrary, both his account of space and time as forms of human sensibility, which leaves open the possibility of alternative forms, and his view of divine knowledge as nonsensible are incompatible with the claim that space and time are "the unique conditions of the representation of objects" (358). Thus, space and time are only conditionally necessary for Kant, since their necessity is limited to beings with our particular forms of sensibility. Moreover, it is precisely because they involve this conditional necessity that Kant claims that they are transcendentally ideal. Guyer's failure to recognize this is not due to his ignorance of the texts, but to his refusal to take seriously the transcendental dimension of Kant's thought. This refusal is a virtual leitmotif of the book; but it finds its clearest expression in a remarkable passage in which he states that a conditional necessity "would seem to be the only kind of necessity that we could ever arrive at by any investigation construed, in Lockean fashion, as an exploration of the limits of our own cognitive faculties —which is to say, by a Copernican revolution as Kant apparently intended that to be understood" (367). If the *Critique of Pure Reason* is seen in this way, it is no wonder that so little of its essential content survives.

HENRY E. ALLISON

University of California/San Diego

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

NOTES AND NEWS

The editors report with sorrow the death of Mark Carl Overvold, Associate Professor of Philosophy and Chair of the Department of Philosophy at Virginia Commonwealth University, on November 2, 1988. Professor Overvold graduated from St. Olaf College in 1970, and received his doctorate from the University of Michigan in 1976. He taught at Brown University, the College of William & Mary, and the University of Texas before joining the faculty at Virginia Commonwealth University in 1977. He was 40 years old at the time of his death. A memorial fund has been established for the education of his children. Contributions may be made to the Mark Overvold Fund, care of James Beran, 2101 California Street, Apt. 104, Mountain View, CA 94040.

The Austrian Ludwig Wittgenstein Society is pleased to announce the Fourteenth International Wittgenstein Symposium on August 13–20 in Kirchberg am Wechsel, Austria. The general title of this centenary celebration is "Wittgenstein: Toward a Reevaluation." Further information may be obtained by writing Rudolf Haller, Institüt für Philosophie, Karl-Franzens-Universität, Heinrichstrasse 26, A-8010 Graz, Austria.

The History and Philosophy of Science in School Science Teaching would like to announce an international conference on November 6–10, sponsored by the Science Education and Department of Philosophy at Florida State University. The conference is designed to provide for interchanges among teachers, teacher educators, school and regional administrators, scientists, science historians, and philosophers of science. Papers on the place of the history and philosophy of science in school science teaching may be submitted in abstract form by June 15. Additional information may be obtained by writing Kenneth Tobin, Science Education, Carrothers Hall, Florida State University, Tallahassee, FL 32306-3032.

The Institute for Philosophy and Public Policy at the University of Maryland will hold its annual summer Workshop for Philosophers on June 21–23 at the Catholic University of America. Further information may be obtained by writing Kathleen Wiersema, Institute for Philosophy and Public Policy, 0123 Woods Hall, University of Maryland, College Park, MD 20742.

The Department of Philosophy at the University of Rochester is pleased to announce the establishment of an annual international competition for the best new and unpublished essay in English on the philosophy of George Berkeley. The deadline for the competition will be November 1 of each year. Submissions will be judged by members of a review board selected by the department at Rochester. The winner shall be announced on March 1 of the following year. Submissions may be sent to Ralf Meerbote, Philosophy Department, University of Rochester, Rochester, NY 14627.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The Department of Philosophy at Rutgers University is pleased to announce the winner of the Second Annual Rutgers International Prize in Philosophy for an outstanding undergraduate paper. He is Matthew Schnall from Harvard College, and his paper is entitled "Assertibility Plus: Reasoning with Conditionals." The deadline for the Third Annual Prize is November 1, 1989.

The Sixteenth Annual St. Louis Conference on Manuscript Studies will be held at St. Louis University on October 13–14. An invitation has been extended for papers dealing with one of the following aspects of manuscript: codicology, illumination, paleography, and texts. Additional information may be obtained by writing Dr. Tolles, Conference Committee, Pius XII Memorial Library, 3650 Lindell Boulevard, St. Louis, MO 63108.

The Hegel Society of America will hold its eleventh annual meeting on October 12–14, 1990, at Bates College. Papers are invited that deal with the interpretation or application of Hegel's Philosophy of Religion. Deadline for submission is January 31, 1990. Decisions will be made by May 15, 1990. Further information may be obtained by writing David Kolb, Philosophy Department, Bates College, Lewiston, ME 04240.

The Society for Health and Human Values is pleased to announce that its 1989 annual meeting will take place on October 26–29 in Washington, DC. The Society welcomes the submission of papers on topics in the humanities and health sciences, including but not restricted to ethics, philosophy, religious studies, literature, history, creative arts, and law. Social science, educational evaluation, and health policy studies related to the humanistic aspects of health care are also welcome. Deadline for receipt of papers is June 15. Papers may be submitted for blind review to the SHHV Annual Meeting Program, 6728 Old McLean Village Drive, McLean, VA 22101. Further information about the meeting may be obtained by writing John Moskop, Department of Medical Humanities, East Carolina University School of Medicine, Greenville, NC 27858-4354.

The London School of Economics and Political Science would like to announce that the third Lakatos Award has been awarded to Michael Redhead of University of Cambridge, for his book, *Incompleteness, Nonlocality and Realism*. The Award, which is for an outstanding contribution to the philosophy of science, has been endowed by the Latsis Foundation in memory of Imre Lakatos. It will be awarded annually for at least ten years.

The General Editor of the critical edition of *The Works of George Santayana* would appreciate receiving relevant first-hand material, such as letters, personal recollections, photographs, unpublished writings, or information concerning such materials. All material shall be treated with the greatest care

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

and promptly returned. Materials may be directed to Herman J. Saatkamp, Philosophy Department, Texas A & M University, College Station, TX 77843-4237.

The Department of Philosophy at the University of Chicago is very pleased to announce that Arnold Davidson has been promoted to Associate Professor with tenure.

The Department of Philosophy at the University of Illinois/Chicago is pleased to announce the appointments of Debra DeBruin, Graciela De-Pierris, and Marya Schechtman as Assistant Professors.

The Department of Philosophy at Wellesley College is pleased to announce the appointment of Alison McIntyre as Assistant Professor.

The Department of Philosophy at the University of Wisconsin/La Crosse is pleased to announce the appointment of Eric R. Kraemer as Associate Professor.

The Journal of Medicine and Philosophy

Sponsored by the Center for Ethics, Medicine, and Public Issues, a joint project of Baylor College of Medicine, the Institute of Religion, and Rice University, in collaboration with the School of Medicine, University of Connecticut Health Center

Editor
H. Tristram Engelhardt, Jr.

Associate Editor
Stuart F. Spicker

The Journal of Medicine and Philosophy continues to explore the shared themes and concerns of philosophy and the medical sciences.

Subscription Information **ISSN** 0360–5310
1988, Volume 13 (4 issues)
Institutional rate: Dfl. 161.00/US$76.00 incl. postage/handling
Private rate: Dfl. 78.00/US$26.00 incl. postage/handling
Single issues: Dfl. 28.00/US$11.50 incl. postage/handling
Special rate for SHHV members: Dfl. 70.00/US$23.00 incl. postage/handling

Private subscriptions should be sent direct to the publishers

D. Reidel Publishing Company

Kluwer Academic Publishers
P.O. Box 989, 3300 AZ Dordrecht, The Netherlands
101 Philip Drive, Norwell, MA 02061, U.S.A.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVI, NO. 5, MAY 1989

ON THE SHARED PREFERENCES OF TWO BAYESIAN DECISION MAKERS*

AN outstanding challenge for "Bayesian" decision theory is to extend its norms of rationality from individuals to groups. Specifically, can the beliefs and values of several Bayesian decision makers be amalgamated into a single Bayesian profile that respects their common preferences over options? If rational parties to a negotiation can agree on collective actions merely by considering mutual gains, is it not possible to find a consensus Bayes model for their choices? In other words, can their shared strict preferences over acts be reproduced with a Bayesian rationale (maximizing expected utility) from beliefs (probabilities) and desires (utilities) that signify a rational compromise between their rival positions?

Whatever else is to be required of a compromise, we suppose that a consensus Bayes model for the group preserves those strict preferences which the individuals already share. That is, we impose a *weak Pareto condition* on compromises. Whenever all parties to a decision have a common strict preference for one option over another, then any proposed Bayesian group model for their choice—any "neutral" position—must reflect this preference and assign higher expected utility to the Pareto dominating option.

Of course, the probabilities and utilities of any one of the agents satisfies this weak Pareto condition. That is, each agent on her own meets this condition—whatever strict preferences they all have, each has. But, it is hardly a compromise always to make the group decide all questions based on the preferences of a single member. We call

* We would like to thank John Broome, Robyn Dawes, Jay Goodman, Mark Kamlet, and Isaac Levi. This work was supported by the Buhl Foundation, ONR Contracts N00014-85-K-0539 and N00014-88-K-0013, and NSF Grant DMS-8805676.

0022-362X/89/8605/225–244 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

such Pareto solutions *autocratic*. What (nonautocratic) Bayesian compromises are there for the group decisions?

Imagine the dilemma that arises when two Bayesian agents agree on what to do by appeal to their unanimous preferences, but they find no neutral Bayes position (other than their own separate views) to endorse the rationality of their choices. It is our primary purpose in this essay to underscore the ubiquitous nature of this dilemma. Our central result, then, is this. When two Bayesian decision makers differ both in their beliefs (probabilities) and values (utilities), and subject to an assumption that they agree on the preference ranking of some two "constant" acts, the only candidates for a Bayes compromise are the two autocratic solutions. That is, there is no room for a Bayesian compromise. We develop this argument in section I.

In section II we contrast our results with several related theorems in social welfare theory. The question of finding a neutral Bayes model for the group's choices is seen at once to be a variety of Kenneth Arrow's[1] problem for social welfare rules. The Bayesian agents have individual preferences over social acts which are to be amalgamated into a single, Bayesian group preference over social acts. In order to dodge that "impossibility" theorem, some concession to Arrow's result is required.

The alternatives that have been investigated in connection with Arrow's problem follow several directions. One approach is to restrict the domain of applicability of a welfare rule to communities where the individual preferences are not too discrepant, e.g., where the individual preferences conform to Duncan Black's[2] "single-peakedness" condition. (This violates the "unrestricted domain" condition of Arrow's argument.) Our negative finding is unaffected by this consideration, since we show that the dilemma arises for all pairs of Bayesian agents who have even the slightest differences in

[1] *Social Choice and Individual Values* (New York: Wiley, 1951). Stated briefly, Arrow's *impossibility theorem* shows that it is inconsistent to posit a social choice rule for amalgamation of individual preferences into a group preference, subject to four conditions: (1) the rule has unrestricted domain—it applies with all varieties of individual preferences and all sets of feasible acts; (2) the rule satisfies the weak Pareto condition—if every one (strictly) prefers option A to option B, then the social ordering makes A strictly better than B; (3) the rule precludes *dictator* solutions—it cannot be that the social ordering is determined by one individual's preferences regardless the preferences of all others; (4) the rule conforms to an *independence of irrelevant alternatives* condition—the social ordering of a set of feasible options depends solely on the individuals' orderings for these options. (The fourth condition combines a prohibition on interpersonal utility comparisons with an assumption that the social choice rule induces an ordering of acts by revealed preferences.)

[2] "The decision of a committee using a special majority," *Econometrica*, XVI (1948): 245–261.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

both their beliefs and their values. There is no gain made here by trying to bracket cases where the agents' probabilities or utilities reflect large, rather than small, discrepancies.

A second approach to avoiding Arrow's impossibility theorem is to add structure to the representation of individual preferences, to add interpersonal utility comparisons not allowed by Arrow's multipartite "independence of irrelevant alternatives" condition. The excellent paper of Kevin Roberts[3] summarizes how different classes of social welfare rules can be achieved by introducing alternative versions of positive interpersonal utility comparisons. Our negative finding about Bayesian social welfare rules applies, however, whether or not interpersonal utility comparisons are made. Thus, another familiar way around Arrow's result does not work when the problem is finding group compromises that are Bayes.

A third approach is to relax the "ordering" requirement, to liberalize the condition that the social choice rule induces a complete ordering of group options where any two social acts may be compared by the compromise social preference relation. In his seminal book on Bayesian decision theory and statistics, Leonard Savage[4] defends the minimax-regret rule for group deliberation by suggesting that the "ordering" postulate ($P1$) does not apply to group preference, though it does apply to individual preferences.

In this connection, Isaac Levi[5] offers an intriguing account, we think, why the norms on rational choice should be uniform between groups and individuals. The key assumption for a group decision problem is that there is only one agent, the (cooperative) group. Otherwise, there is not one decision problem but, instead, there are the several (noncooperative) deliberations of the separate individuals who, for prudential reasons, take an interest in each other's actions.

In section II.2 we explain how our result impacts on Levi's proposal for achieving a unified decision theory, unified across individual and group decisions. Levi's theory relaxes the "ordering" postulate of Bayesian expected utility. A rational agent (either an individual or a group) need not have a complete preference relation for comparing every two options. The explanation for this departure from Bayesian theory is his view that a rational agent (either an individual or a group) may experience *unresolved* conflicts.

For an individual, the conflicted preferences arise from uncertain beliefs (sets of probabilities) or from indeterminate values (sets of

[3] "Possibility Theorems with Interpersonally Comparable Welfare Levels," *Review of Economic Studies*, XLVII (1980): 409–420.
[4] *The Foundations of Statistics* (New York: Wiley, 1954), §13.5.
[5] "Conflict and Social Agency," this JOURNAL, LXXIX, 5 (May 1982): 231–247.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

utilities). For a group, the conflicts among the individuals' beliefs and values generate the same uncertainties and indeterminacies when the agent is the cooperative group.

In Levi's theory, however, the compromise position on rational choice is achieved (at the expense of "ordering" for the preference relation) by constructing two *independent* "neutral" positions, one for conflicted beliefs (a set of probabilities) and one for conflicted values (a set of utilities). An incomplete preference relation is formed using inequalities in expected utilities with all pairs from these two sets. Based on our negative result, we show this rule leads to violations of the weak Pareto condition. That is, Levi's theory allows a group to choose a weak Pareto dominated option.

Nonetheless, we are in agreement with Levi on the desirability of relaxing the ordering postulate while preserving a respect for expected utility. Hence, we explore a theory in which preference need not be a complete relation, but also a theory in which preference does not compromise conflicts in beliefs and in values independently.

I. BAYESIAN COMPROMISES BETWEEN TWO BAYESIANS

Here we discuss in detail the case of a group composed of just two Bayesian decision makers. That is, we respond to the question posed in the opening sentences of this essay when only two Bayesian "experts" are involved and, for each, preference is a complete relation. We assume that each agent can compare every two acts, either by a (transitive, antisymmetric) relation of strict preference, or else by an equivalence relation of indifference.

The group decision involves amalgamating the preferences of two Bayesians, whom we shall call Dick and Jane. According to Bayesian decision theory, for simple problems, an agent chooses from among a set of feasible (state independent[6]) acts according to the principle of maximizing subjective expected utility.

[6] We assume acts and states are probabilistically independent. Later on where we use horse lotteries for acts, as explained in fn. 9, we assume also that states are value-neutral with respect to the lottery prizes.

Interesting discussions of the measurement problem for expected utility theory without either of these assumptions is found in J. Drèze, "Decision Theory with Moral Hazard and State–Dependent Preference," paper #8545, *C.O.R.E.* (1985), Universite Catholique de Louvain, Voie du Roman Pays, 34, B-1345, Louvain–la-Neuve, Belgium.

For additional, important commentary about the effects of state-dependent utilities on the measurement of probabilities, see E. Karni, D. Schmeidler, and K. Vind, "On State Dependent Preferences and Subjective Probabilities," *Econometrica*, LI, 4 (July 1983): 1021–1031; H. Rubin, "A Weak System of Axioms for 'Rational' Behavior and the Non-Separability of Utility from Prior," *Statistics and Decisions*, v (1987): 47–58; and J. B. Kadane and R. L. Winkler, "Separating Probability Elicitation from Utilities," *Journal American Statistical Association*, LXXXIII, 402 (1988): 357–363.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

To be more precise, consider the familiar decision matrix:

Decision Matrix: Acts **X** States

	$S1$	$S2$			Sj			Sn
A_1	O_{11}	O_{12}			O_{1j}			O_{1n}
A_2	O_{21}	O_{22}			O_{2j}			O_{2n}
A_m	O_{m1}	O_{m2}			O_{mj}			O_{mn}

The columns denote a partition into (n) states of nature, $S_1, \ldots, S_n$, about which the agent is uncertain. The rows designate feasible acts, $A_1, \ldots, A_m$, whose outcomes in each state are denoted by the O_{ij}: the outcome of act A_i under state S_j. The agent's uncertainties about the states are given by a probability distribution, $P(S_j)$. The agent's values for outcomes are given by a (von Neumann-Morgenstern) utility function, $U(O_{ij})$. Then, according to the principle of maximizing expected utility, act A_1 is (strictly) preferred to act A_2 whenever

$$\sum_j P(S_j) \cdot U(O_{1j}) > \sum_j P(S_j) \cdot U(O_{2j}).$$

Suppose Dick's preferences over such acts are summarized by the pair (P_1, U_1) of his (personal) probability and utility, while Jane's preferences are depicted by the pair (P_2, U_2). What are the alternative Bayesian preference schemes that agree with the (strict) preferences shared by Dick and Jane? That is, for which pairs (P, U) is it the case that:

$$\sum_j P(S_j) \cdot U(O_{1j}) > \sum_j P(S_j) \cdot U(O_{2j})$$

whenever

$$\sum_j P_k(S_j) \cdot U_k(O_{1j}) > \sum_j P_k(S_j) \cdot U_k(O_{2j}) \qquad (k = 1, 2)?$$

In terms of social welfare rules, our question is this. In group choices where all agents receive the same outcomes (the social acts

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

offer the same prospects to each agent), subject to a weak Pareto condition, what are the Bayesian social welfare rules? (Recall, a social choice rule satisfies the weak Pareto condition provided that an option is inadmissible whenever there is another feasible alternative that everyone strictly prefers. If everyone strictly prefers option B to option A, A is inadmissible whenever B is available.)

The answer is surprising. Stated informally, our result in the case of two agents is that no attractive compromises exist. Only autocratic solutions conform to the weak Pareto condition. That is, a Bayesian model for the group's collective decisions must use the beliefs and values of a single agent, thereby ignoring the preferences of everyone else whenever the weak Pareto condition does not apply.[7]

It is interesting to contrast the difference between the weak and strong Pareto conditions for group compromises. The strong Pareto condition for social choice rules requires that option A is socially inadmissible whenever there is a feasible option B, which everyone finds either strictly preferable or indifferent to A, and which someone strictly prefers to A. We find that with the imposition of a strong Pareto condition, there are *no* Bayes social welfare rules.

The proofs of these claims are contained in the following, simple example. The constructions in the example generalize to every pair, like Dick and Jane, that differs in both beliefs and values.

Example 1: Suppose, as before, Dick and Jane are Bayesian agents with preferences summarized by the probability and utility pairs (P_k, U_k), $k = 1, 2$. Assume that they have different beliefs, $P_1 \neq P_2$. That is, there is some event E with $P_1(E) \neq P_2(E)$. For instance, let Dick assign E a (personal) probability .1 while Jane assigns E a (personal) probability .3, i.e., $P_1(E) = .1$ and $P_2(E) = .3$. Also, suppose they have different values, $U_1 \neq U_2$.[8] For simplicity, suppose Dick and Jane

[7] This is not the same as Arrow's concept of a "dictator," since the determination of who is the autocrat may be a function of everyone's preferences, contrary to the requirements for an Arrovian dictator.

[8] This assumption is complicated by controversies over interpersonal utility comparisons. With or without concern for another's preferences, an individual's utility function is defined up to an affine transformation. Regarding this, we suppose that for any two agents there is some pair of v.N–M lotteries, L_* and L^*, which are ranked the same by both—each (strictly) prefers L^* to L_*. (This is mild, as we may introduce two new rewards with this property. We discuss the case of completely opposed preferences in fn. 11.) In case interpersonal utilities are recognized, consider the (perhaps, two) representation(s) of the agents' joint utilities:

$$U_a = \{U_{1a}, U_{2a}\} \text{ and } U_b = \{U_{1b}, U_{2b}\},$$

where

$$U_{1a}(L_*) = U_{2b}(L_*) = 0 \text{ and } U_{1a}(L^*) = U_{2b}(L^*) = 1.$$

Then, the assumption that Dick and Jane have different values is expressed by saying that there is a lottery, L, with $U_{1a}(L) \neq U_{2b}(L)$. That is, their (separate) utility functions do not coincide under affine transformations. Since we are concerned

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Table 1: "Horse lotteries" used in fixing the upper and lower probabilities and utilities.

	E	$\neg E$
	r^*	r_*
A_1	$(.9 + \epsilon)(r_*) + (.1 - \epsilon)(r^*)$	$(.9 + \epsilon)(r_*) + (.1 - \epsilon)(r^*)$
$A_{2\epsilon}$	$(.7 - \epsilon)(r_*) + (.3 + \epsilon)(r^*)$	$(.7 - \epsilon)(r_*) + (.3 + \epsilon)(r^*)$
$A_{3\epsilon}$	r	r
A_4	$(.6 - \epsilon)(r_*) + (.4 + \epsilon)(r^*)$	$(.6 - \epsilon)(r_*) + (.4 + \epsilon)(r^*)$
$A_{5\epsilon}$		

have different (cardinal) utilities, in the following sense. They agree on the ranking of two particular rewards: each prefers r^* to r_*, though they differ in their valuation of a third reward r. Let r be a reward with $U_1(r) = .1$ and $U_2(r) = .4$, while there exist rewards r_* and r^* with $U_1(r_*) = U_2(r_*) = 0$ and $U_1(r^*) = U_2(r^*) = 1$.

Next, we consider pairs of options where Dick and Jane hold common preferences. Examine the acts defined in table 1, above, on the binary partition formed by the event E.

These acts are "horse lotteries," in the language of F. J. Anscombe and R. J. Aumann.[9] We distinguish a von Neumann–Morgenstern (v.N–M) lottery from a horse lottery. A v.N–M lottery, L, is a specified probability distribution over a set of rewards. For example, a v.N–M lottery that yields the reward r_* with probability .6 and the reward r^* with probability .4 is denoted by: $L = .4r_* + .6r^*$. With horse lotteries as acts, the outcome O_{ij} of act A_1 in state s_j (see the decision matrix on page 229) is a v.N–M lottery L_{ij}.

---

with the set of Bayes compromises that preserve unanimous (strict) preferences, we may use the pair (U_{1a}, U_{2b}) to represent the agents' individual utilities while also respecting interpersonal utility comparisons. For convenience, abbreviate U_{1a} by U_1 and U_{2b} by U_2.

[9] "A definition of subjective probability," *Annals of Mathematical Statistics*, XXXIV (1963): 199–205. Their theory of "horse lotteries" is summarized by four axioms on preference. Informally, these four axioms require that: (A1) preference is a weak order; (A2) preference satisfies the "independence" condition (related to Savage's "sure-thing" postulate *P*2); (A3) preference obeys an Archimedean condition—to insure utilities are real-valued; and (A4) preference for v.N–M lotteries is state-independent.

Specifically, the final axiom requires this. Let H_1 be the "constant" horse lottery that awards the same v.N–M lottery L_1 in each state and let H_2 be the "constant" horse lottery that awards L_2 in each state. Let horse lotteries $H_{1'}$ and $H_{2'}$ differ solely in that, for some state s, $H_{1'}(s) = L_1$ and $H_{2'}(s) = L_2$. $H_{1'}$ and $H_{2'}$ have the same outcomes in every other state. Then axiom A4 says: H_1 is preferred to H_2 if and only if $H_{1'}$ is preferred to $H_{2'}$.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

For example, the v.N–M lottery $(.9 + \epsilon)(r_*) + (.1 - \epsilon)(r^*)$ under event E for act $A_{2\epsilon}$ corresponds to the outcome: receive reward r_* with probability $(.9 + \epsilon)$ and reward r^* with probability $(.1 - \epsilon)$. When "horse$_j$" wins the hypothetical race, i.e., when state s_j obtains, the act A_i pays out the v.N–M lottery L_{ij}. Thus, horse lotteries are functions from states to v.N–M lotteries.

Since which state obtains (which "horse" wins) is uncertain, and how rewards are to be valued is not stipulated in advance, "horse lotteries" accommodate both uncertainty over states and (cardinal) utility over rewards.

First, as Dick and Jane agree that

$$.1 \leq P_k(E) \leq .3 \qquad (k = 1, 2)$$

they agree that act A_1 is preferred to each act $A_{2\epsilon}$, and they agree that each act $A_{3\epsilon}$ is preferred to A_1 $(.1 \geq \epsilon > 0)$. See table 1.

Likewise, they agree that

$$.1 \leq U_k(r) \leq .4 \qquad (k = 1, 2).$$

Then, also they are unanimous in their preference for A_4 over each act $A_{2\epsilon}$, while each act $A_{5\epsilon}$ is preferred to A_4. (All these preferences are "strict.")

In figure 1 we see the set of pairs of probabilities and utilities agreeing with these shared preferences. That is, figure 1 is based on the (strict) preferences involving the upper and lower probabilities of event E and the upper and lower utilities of reward r. These bounds for $P(E)$ and $U(r)$ box the family of Bayesian compromises between Dick and Jane with respect to their shared agreements for these (strict) preferences. Subject to the weak Pareto condition, it shows that the set of "neutral" Bayes' models (with respect to *all* their choices—not just for these few comparisons) is some subset of the cross product of weighted averages of their probabilities and weighted averages of their utilities.

Next, consider the set of pairs of acts defined in table 2, over which Dick and Jane also hold common preferences.[10]

For each value, $0 < \epsilon \leq .2$, Dick and Jane agree in their (strict) preference for $A_{7\epsilon}$ over $A_{6\epsilon}$. The set of Bayesian compromises for these preferences is *not* connected, however. See figure 2.

For $0 < \epsilon < .015$ the set of probability/utility pairs $(P(E), U(r))$ for which $A_{7\epsilon}$ has greater expected utility than $A_{6\epsilon}$, is bounded by a hyperbola centered at (.2, .25), which satisfies:

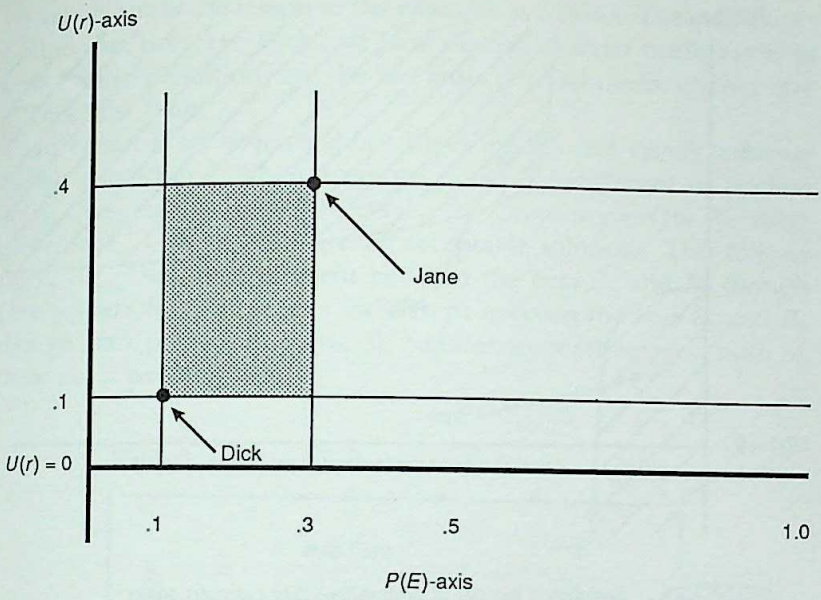
$$[P(E) - .2][U(r) - .25] = .015 - \epsilon.$$

[10] The particularly convenient form of these horse lotteries is due to Jay Goodman, formerly of the Statistics Dept. at Carnegie Mellon University.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Figure 1



designates the set of probability/utility pairs agreeing with the common preferences of Dick and Jane for the comparisons, above, in table 1.

As $\epsilon \Rightarrow 0$, the hyperbola approaches the pair of points corresponding to Dick and Jane's preferences. When $\epsilon = 0$, the hyperbola intersects these two points.

If we superimpose the two figures, we obtain figure 3.

Figure 3 shows that the family of Bayesian agents who agree with these two, even for the few preferences already considered, consists exactly of the two themselves: (P_1, U_1) and (P_2, U_2). That is, since the hyperbolas all have negative slopes at the points interior to the box, as $\epsilon \Rightarrow 0$ the regions of overlap between the two figures collapse onto

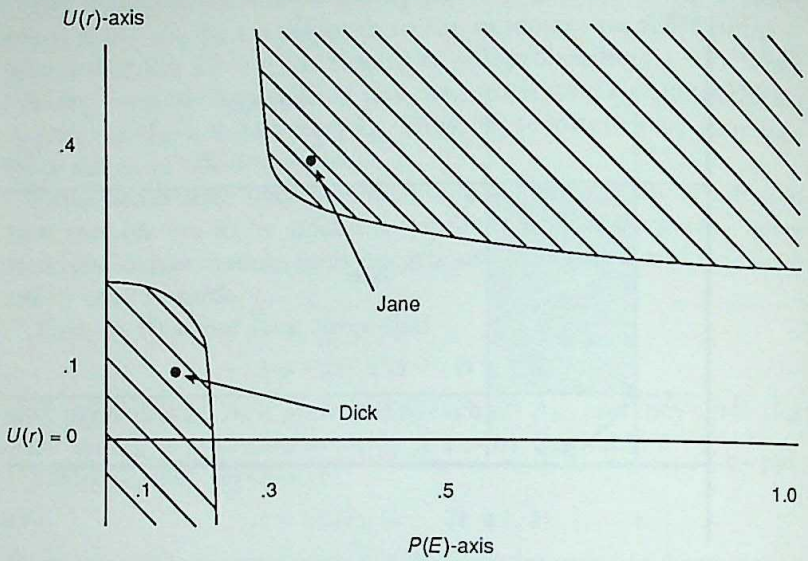
Table 2: "Horse lotteries" used in separating the set of compromises between Dick and Jane.

	E	$\neg E$
$A_{6\epsilon}$	$.785(r_*) + .215(r^*)$	$\epsilon(r_*) + .2(r) + (.8 - \epsilon)(r^*)$
$A_{7\epsilon}$	$(.2 - \epsilon)(r_*) + .8(r) + \epsilon(r^*)$	$.165(r_*) + .835(r^*)$

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

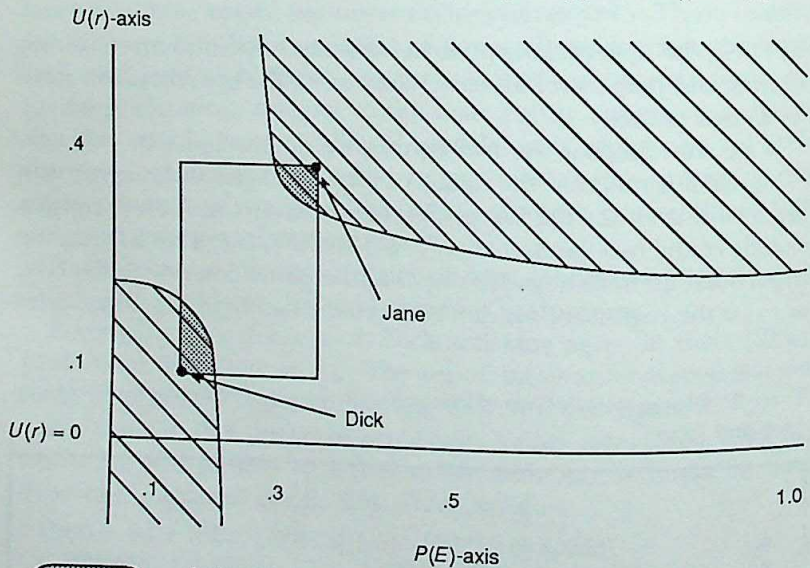
 Digitized by Arya Samaj Foundation Chennai and eGangotri

Figure 2: Preferences which separate the family of agreeing probability/utility pairs.



designates the set of probability/utility pairs agreeing with the common preferences of Dick and Jane for $A_{7\varepsilon}$ over $A_{6\varepsilon}$, $\varepsilon = .01$, in table 2.

Figure 3: Preferences which separate the family of agreeing probability/utility pairs.



designates the set of probability/utility pairs agreeing with the common preferences of Dick and Jane, $\varepsilon = .01$, in tables 1 and 2.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the two corner points: the points corresponding to Dick and Jane. We may express the lesson of the example as follows. The only Bayes models that preserve Dick and Jane's common strict preferences (a weak Pareto condition) are the two autocratic solutions: choose one of Dick and Jane.

Moreover, if we inquire about Bayes models that satisfy a strong Pareto condition ("When everyone is either indifferent or prefers act A_1 to A_2 and someone prefers A_1, the compromise is for the strict preference."), then there are *no* acceptable solutions. This follows since, e.g., Dick is indifferent between the bets B_1 and B_2 though Jane prefers B_1, and Jane is indifferent between the bets B_1 and B_3 though Dick prefers B_3 (table 3). Neither autocrat respects both of these strict preferences.

Table 3: Bets used with the strong Pareto condition.

	E	$-E$
B_1	r^*	r_*
B_2	$.9(r_*) + .1(r^*)$	$.9(r_*) + .1(r^*)$
B_3	$.7(r_*) + .3(r^*)$	$.7(r_*) + .3(r^*)$.

The various constructions in this example generalize easily to any pair of agents with different beliefs and values, subject to the common ranking for some pair of v.N–M lotteries, L^* and L_*.[11] (We choose $U_k(L_*) = 0$ and $U_k(L^*) = 1$, $k = 1, 2$.)

[11] This assumption cannot be removed. A result due to J. Kadane ["Opposition of Interest in Subjective Bayesian Theory," *Management Science,* XXXI (1985): 1586–1588, Theorem 1] says that two agents hold opposing strict preferences over all pairs of acts if, and only if, they share a common personal probability for the states and have opposite (cardinal) utilities for the rewards. Then, as a simple corollary, if two agents hold different probabilities and diametrically opposed (cardinal) utilities, there will be some pairs of horse lotteries which, by strict preference, they rank in common.

That is, suppose Susan and John have different personal probabilities (denoted P_S and P_J, with $P_S \neq P_J$) and strictly opposed preferences for all pairs of v.N–M lotteries, the "constant" horse-lottery acts. Thus, whenever John strictly prefers lottery L_1 to lottery L_2 Susan's preference goes the other way, and vice versa. There is no pair of lotteries which they rank order the same. Then, after standardizing their separate utility units with a single pair of v.N–M lotteries (as in fn. 8), where Susan prefers the first and, hence, where John prefers the second of the two lotteries, we have: $U_S = 1 - U_J$.

Nonetheless, as $P_S \neq P_J$, by Kadane's result, there are pairs of horse lotteries which they rank order in the same way. Let E be an event with

$$P_S(E) = p_S < p_J = P_J(E).$$

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Theorem 1: Let (P_1, U_1) and (P_2, U_2) be the probability/utility pairs representing two decision makers with different beliefs and preferences over horse lotteries, then:

(i) The set of probability/utility pairs which agree with the strict preferences shared by these two decision makers (weak Pareto) consists exactly of the two pairs themselves—there are autocratic solutions only—no other Bayesian compromises exist; and

(ii) Subject to the strong Pareto condition, the set of probability/utility pairs that agree with these common preferences is empty—there are no Bayes models at all.

Proof: Since $P_1 \neq P_2$ and $U_1 \neq U_2$, there is an event E and a v.N–M lottery L such that $P_1(E) = p_1 < P_2(E) = p_2$, and $U_1(L) = u_1 < U_2(L) = u_2$. Then the agents agree that:

$$p_1 \leq P_k(E) \leq p_2$$

$$u_1 \leq U_k(L) \leq u_2 \qquad (k = 1, 2).$$

(*i*) Table 1 is modified by these bounds, creating a "box of compromises," as in figure 1. The pairs of acts used to "separate" the set of compromises, analogous to those of table 2, are defined by the hyperbolic equations (provided by Jay Goodman):

$$(P(E) - [p_1 + p_2]/2) \cdot (U(E) - [u_1 + u_2]/2) = (p_2 - p_1)(u_2 - u_1)/4 - \epsilon$$

for $0 < \epsilon < (p_2 - p_1)(u_2 - u_1)/4$. Combining these sets of preferences, letting $\epsilon \Rightarrow 0$, we discover that only the original pairs (P_1, U_1) and (P_2, U_2) remain. By holding fixed the binary partition $\{E, \neg E\}$ and varying the lottery L, we see that the group utility function must be the utility for one of the two agents. Then, by varying the partition, since utility is state independent, we see that the group probability also must be taken from the same agent.[12] Hence, only autocratic solutions agree with the weak Pareto condition on strict preferences.

Consider two horse lotteries $H_1 = \{L_{11}, L_{12}\}$ and $H_2 = \{L_{21}, L_{22}\}$, defined on the binary partition $\{E, \neg E\}$, where the L_{ij} $(i, j = 1, 2)$ are v.N–M lotteries. John prefers H_1 to H_2 if, and only if, $cp_J + d > 0$, where
$c = [U_J(L_{11}) + U_J(L_{22}) - U_J(L_{12}) - U_J(L_{21})]$
and $d = [U_J(L_{12}) - U_J(L_{22})]$. Because $U_S = 1 - U_J$, Susan also prefers H_1 to H_2 if, and only if, $cp_S + d < 0$. Hence, with $d < 0 < c$ and $-p_J c < d < -p_S c$, John and Susan have common strict preferences for H_1 over H_2.

Other Bayesians will agree with John and Susan in preferring H_1 to H_2 provided they have preferences of the form: (P^*, U_J) or (P_*, U_S) where $P^*(E) > P_J(E)$ and $P_S(E) > P_*(E)$ whenever $P_J(E) > P_S(E)$. That is, there exist nonautocratic, Bayes models for the set of weak Pareto agreements between Susan and John. These models correspond to Bayesian agents with *more extreme* degrees of belief (combined with the respective utility)—models that exaggerate the expected utility differences between the two of them. None of these more extreme Bayes models strikes us as a serious compromise between Susan and John. Thus, even in this exceptional case of diametrically opposed utilities, we cannot locate a viable Bayes compromise for their unanimous preferences.

[12] The proof depends upon the "state independence" of utilities. Without it, i.e., if we require only axioms A1–A3 (see fn. 9) there are a continuum of (weak) Pareto,

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(*ii*) Upon adding a strong Pareto requirement for compromises, we may modify the bets from table 3, to read as follows:

Modified Table 3: Bets used with the strong Pareto condition.

	E	$\neg E$
	L^*	L_*
B_1	$(1-p_1)(L_*)+p_1(L^*)$	$(1-p_1)(L_*)+p_1(L^*)$
B_2	$(1-p_2)(L_*)+p_2(L^*)$	$(1-p_2)(L_*)+p_2(L^*)$.
B_3		

Then, as before, the first agent is indifferent between B_1 and B_2 though the second prefers B_1, and the second agent is indifferent between B_1 and B_3 though the first prefers B_3. Neither "autocrat" respects both these strict preferences. Hence, under a strong Pareto condition on compromises, there are none. □

II. GROUP DECISION MAKING AND A SEPARATION OF BELIEF AND UTILITY

II.1. Social Welfare Theory: An immediate consequence of theorem 1 is the impossibility of a general, nonautocratic Bayesian social welfare function (subject to the weak Pareto condition). That is, even when interpersonal utility comparisons are admitted, and provided the domain of social acts include the simple varieties discussed in the previous section, there is no interesting Bayesian solution to the social welfare problem.

That corollary to theorem 1 is obvious when the population consists in two agents. For larger communities, of size n, consider populations where the beliefs and preferences of different agents fall into one of two camps, i.e., where the n-agents are clones of two agents.[13] Thus, the requirement that group welfare satisfies the strong Pareto condition (or else a conjunction of "weak Pareto" and "no autocrats") is inconsistent with the rationality postulates of subjective

---

Bayesian compromises using state-dependent utilities. This follows from two other results: (1) Theorem 13.1, p. 176, of P. Fishburn's *Utility Theory for Decision Making* (New York: Krieger, 1979); and (2) The existence of a (convex) set of preferences, satisfying axioms A1, A2, and A3, each of which extends the strict partial order formed by the (weak) Pareto condition. Result (2) is proven in our unpublished technical report, "A Representation for Preference as a Strict Partial Order in Terms of Sets of Pairs of Probabilities and Utilities," in preparation, Dept. of Statistics, Carnegie Mellon Univ.

[13] This mimics the technique used in the two papers we discuss, by A. Hylland and R. Zeckhauser, "The Impossibility of Bayesian Group Decisions with Separate Aggregation of Beliefs and Values," *Econometrica,* XLVII (1979): 1321–1336, p. 1330, equation 7; and by P. Hammond, "Ex-*ante* and Ex-*post* Welfare Optimality under Uncertainty," *Economica,* XLVIII (1981): 235–250, p. 241, for larger communities of n-many decision makers.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

expected utility theory. Moreover, this result is resistant to the standard cures for Arrow's impossibility theorem. Neither considerations of interpersonal utility nor restrictions on the extent of the discrepancies between the beliefs and values of two Bayesians avoids the dilemma.

The impossibility of Bayes solutions to group decisions involving uncertainty is discussed also in two important papers, one by A. Hylland and R. Zeckhauser, and one by P. Hammond.[14] We detail a contrast between theorem 1 and these earlier results in an appendix. Briefly stated, those accounts are limited in two ways that leave theorem 1 unaffected.

(1) The earlier findings establish the impossibility of a general, Bayesian welfare rule subject to the condition that group preference amalgamates probability and utility independently. In other words, they require that the Bayes model for the group has the group probability defined solely in terms of the individuals' probabilities, and has the group utility defined solely in terms of the individuals' utilities. Theorem 1 applies without this limitation.

(2) The earlier findings show that with each (Bayesian) social welfare rule for amalgamation of individual preferences into a group preference (and subject to the independent amalgamation of belief and desire discussed in the previous paragraph), *there is some* profile of individual preferences leading to a failure of the (strong) Pareto condition. Theorem 1, however, shows a failure of the (strong) Pareto condition with each rule and for *every* pair of agents (who differ in both beliefs and desires). Thus, theorem 1 derives a universal statement where the other derives an existential.

II.2. A comparison with Levi's quasi-Bayesian decision theory and a problem of independent compromises of beliefs and desires: In important papers and books, Levi advocates a unified theory of rational decision making under unresolved conflict, unified between individual and group decisions.[15] Both for individuals and groups,

[14] Cf. fn. 13. and see John Broome, "Utilitarianism and Expected Utility," this JOURNAL, LXXXIV, 8 (August 1987): 405–422, who comments on the relationship of Hammond's theorem with utilitartianism. Also, he discusses the problem in two unpublished essays, "Bolker-Jeffrey Decision Theory and Axiomatic Utilitarianism," and "Should Social Preferences Be Consistent?" For a survey of some related issues, see C. Genest and J. Zidek, "Combining Probability Distributions: A Critique and an Annotated Bibliography," with discussion, *Statistical Science*, I (1986): 114–148.

[15] A selection of these works are his "On Indeterminate Probabilities," this JOURNAL, LXXI, 13 (July 18, 1974): 391–418; *The Enterprise of Knowledge* (Cambridge MIT, 1980); "Conflict and Social Agency," *op. cit.*; several entries in *Decisions and Revisions* (New York: Cambridge, 1984), and in *Hard Choices* (New York: Cambridge, 1986).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

there is no requirement that an agent's preferences induce an ordering of options—not all acts need be compared by preference.

Levi's decision theory is a liberalization of (Bayesian) expected utility theory. In his theory, an agent's beliefs are represented by a convex set of (personal) probabilities, $\mathcal{P}$, and preferences for outcomes are represented by a convex set of (cardinal) utilities, $\mathcal{U}$. In brief, an option o is an admissible choice from a set of feasible options $\mathcal{O}$, provided that it satisfies a (lexicographically ordered) sequence of maximizations. The first of these is E-admissibility, which requires that an option maximize expected utility for some probability/utility pair (P, U), where $P \in \mathcal{P}$ and $U \in \mathcal{U}$.

Definition: o is *E-admissible* if and only if

$$\exists(P \in \mathcal{P},\ U \in \mathcal{U})\forall(o' \in \mathcal{O})\ E_{P,U}(o) \geq E_{P,U}(o').$$

A secondary decision test for narrowing the set of E-admissible options is maximizing a "security" index among those options which are E-admissible. Illustrations of "security" include (a) a vacuous standard—all options have equal security; (b) security indexed by worst ($\inf[\mathcal{U}]$) payoff; and (c) security indexed by least ($\inf[\mathcal{P} \times \mathcal{U}]$) expected utility. In addition to these two, Levi[16] entertains ternary, etc., maximization requirements, reflecting added structure in the agent's system of values.

Because of the first condition, E-admissibility, an admissible option is "Bayes," i.e., maximizes expected utility for some probability/utility pair. Hence, admissibility takes expected utility theory as a special case, when both $\mathcal{P}$ and $\mathcal{U}$ are unit sets and, e.g., all other value considerations are vacuous. Even when other value considerations are vacuous, however, if either $\mathcal{P}$ or $\mathcal{U}$ is not a singleton, admissibility fails to induce an ordering.[17]

A central theme in Levi's account of choice under unresolved conflict is that a "neutral" position among conflicting beliefs and desires is a position that preserves the shared agreements between the rivals, yet introduces no judgments over which there is disagreement. If an agent experiences a value conflict between, e.g., two (cardinal) utilities U_1 and U_2 then, according to Levi's theory, the convex combinations of these two,

$$\mathcal{U} = \{\alpha U_1 + (1 - \alpha)U_2 \colon 0 \leq \alpha \leq 1\},$$

[16] *Hard Choices*, §5.7.

[17] Then admissible choices fail to satisfy A. K. Sen's property γ for choice rules, "Social Choice Theory: A Re-examination," *Econometrica*, XLV (1977): 53–89. Hence, it is not even a *normal* choice rule in Sen's terminology. For details, see Seidenfeld's Discussion of A. P. Dempster "Probability, Evidence, and Judgment," in *Bayesian Statistics 2*, Bernardo, DeGroot, Lindely, and Smith, eds. (Amsterdam: North-Holland, 1985), pp. 127–129.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

represents the consensus position of unresolved value conflict from which he makes rational decisions. For example, if John is conflicted between his value assessments of options to acquire art objects—the options rank differently under comparisons of their economic outcomes (U_1) and their aesthetic outcomes (U_2)—nonetheless, by representing his conflicted values with the set $\mathcal{U}$, he may proceed to make rational decisions without first having to resolve the value conflict. The analysis for a consensus position with conflicted beliefs is similar. The "neutral" position for belief is the (convex) set $\mathcal{P}$ of conflicted probabilities.

As we remarked in our introduction, Levi[18] offers arguments that rational deliberation should follow the same standards regardless whether the decision is by an individual or by a cooperative group. The intrapersonal conflicts of values and beliefs for individual decision making should be treated in a like fashion with the parallel interpersonal conflicts for group decisions. Our understanding of his theory is that the "neutral" position for decision under unresolved conflict is found by analyzing desires and beliefs *independently*. That is, the consensus for conflicted preferences over outcomes is a (convex) set of utilities, the consensus for conflicted beliefs is a (convex) set of probabilities, and the independence between them is built into E-admissibility with the appeal to the Cartesian product of these two sets.

If we apply his method to the group decisions faced by Dick and Jane, our example 1, we see that the E-admissible social acts are those which maximize expected utility for some probability/utility pair in the rectangle pictured in figure 1. Then, both acts $A_{6\epsilon}$ and $A_{7\epsilon}$ are E-admissible in a pairwise choice between them. (The expected utility of act $A_{6\epsilon}$ is greater for each probability/utility pair within the unshaded region of the rectangle, depicted in figure 3.) This challenges the claim to the "neutrality" of the E-admissible options, we believe, since E-admissibility fails to respect the shared strict preference for $A_{7\epsilon}$ over $A_{6\epsilon}$ E-admissibility conflicts with the weak Pareto condition.

To emphasize this point, Levi's method of group decision making makes identical the E-admissible options in the feasible sets $\mathcal{O}_\epsilon = \{A_{6\epsilon}, A_{7\epsilon}\}$ for the following three groups of Bayesians. Group_1 consists of Dick and Jane. Group_2 consists of Tom and Mary; where Tom's preferences are summarized by the probability/utility pair (P_1, U_2) and Mary's by (P_2, U_1). Group_3 is composed of all four agents: Tom, Dick, Jane, and Mary. For each of the three groups, both options in the sets $\mathcal{O}_\epsilon$ are E-admissible. Group_1 declares a unani-

[18] "Conflict and Social Agency," *op. cit.*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

mous strict preference for $A_{7\epsilon}$ over $A_{6\epsilon}$, however. Group_2 declares the opposite. And the members of Group_3, of course, find no common preferences over these choices.

Levi's E-admissibility is a normative theory of consensus in *reasons* for preference, the rational causes of decisions. It treats beliefs and values for outcomes as the independent springs for our rational actions. But we discover that this account of consensus in reasons is at odds with the conservation of shared (strict) preferences among options. Then Pareto agreements are taken to be superficial unless they are supported from below by consensus in reasons. Levi's theory makes it a serious question between two exclusive strategies facing group_1: either (merely) appeal to the existing agreements on what to do—choose $A_{7\epsilon}$ without giving reasons or, instead, agree first on the "neutral" reasons for the group's choices and let that consensus dictate admissibility—in which case both options are admissible. How is this higher-order decision problem resolved?

At the expense of denying an independent consensus for conflicts of beliefs and values, we avoid this dilemma. In "Decisions without Ordering,"[19] we propose a theory of choice in which preference over pairs of horse lotteries is a strict partial order. In our theory, preference is represented by a set $\mathcal{S}$ of pairs of probabilities and utilities, as follows. Option o_1 is preferred to option o_2 according to the partial order if and only if the expected utility of o_1 is greater than that of o_2 for each probability/utility pair in $\mathcal{S}$.[20] When the strict partial order characterizes a consensus among different (Bayesian) agents, we take the elements of $\mathcal{S}$ to be all the probability/utility pairs that agree with the shared (weak Pareto) preferences of those agents. Trivially, the weak Pareto condition is respected with this consensus set. Each element of $\mathcal{S}$, each potential compromise among the agents, preserves their shared agreements. Levi's theory does not satisfy this condition for potential compromises.

The sets $\mathcal{S}$ are "conditionally convex": If pairs (P_1, U_1) and (P_1, U_2) belong to $\mathcal{S}$, so too do all pairs of the form (P_1, U_3), where $U_3 = \beta U_1 + (1 - \beta)U_2$ $(0 \leq \beta \leq 1)$. Likewise $\mathcal{S}$ is closed under mixtures of pairs that share a common utility. $\mathcal{S}$ need not be a (convex) cross product of a set of probabilities and a set of utilities, however, as required in Levi's theory. Nor need $\mathcal{S}$ be convex, or even

[19] In W. Sieg, ed., *Acting and Reflecting* (Dordrecht: Reidel, in press). Also available as Tech. Report #391 (1987), Dept. of Statistics, Carnegie Mellon University.

[20] If the weak Pareto condition fails to specify action, i.e., where there are conflicted recommendations based on the expectation inequalities taken from $\mathcal{S}$, then one possibility is Levi's proposal to deploy second-tier considerations in order to choose among those options admissible at the first tier. For example, one might use "security" to compromise choice among the "Pareto" admissible alternatives.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

connected—as illustrated in figure 3. According to this account, the shared preferences for the three groups (above) are represented by three different consensus sets:

$$\mathcal{S}_1 = \{(P_1, U_1), (P_2, U_2)\}$$

$$\mathcal{S}_2 = \{(P_1, U_2), (P_2, U_1)\}$$

$$\mathcal{S}_3 = \{(P, U)\colon P = \alpha P_1 + (1 - \alpha)P_2,\ U = \beta U_1 + (1 - \beta)U_2,\ 0 \leq \alpha, \beta \leq 1\}.$$

Only the third group has (partially ordered) strict preferences agreeing with Levi's E-admissible choices, since $\mathcal{S}_3$ is the cross product of two (convex) sets of probabilities and utilities. In general, the consensus set $\mathcal{S}$ is *not* arrived at by analyzing individual beliefs and values independently.

In another study,[21] we explore a representation for strict partial orders over pairs of horse lotteries in terms of such sets $\mathcal{S}$. It remains for us an open question, in general, what set $\mathcal{S}$ is generated this way by the preferences of several (quasi-) Bayesian agents—agents each of whose preferences is given by a strict partial order of this very sort. With two Bayesian agents, the quasi-Bayesian group preference generated by the weak Pareto condition is reported in theorem 1. Then, $\mathcal{S}$ is the set consisting of the two points. What strict partial order corresponds in this way to the shared preferences of n-Bayesians? To know the answer is to know the Pareto decisions of a panel of Bayesian experts.[22]

TEDDY SEIDENFELD, JOSEPH B. KADANE, and MARK J. SCHERVISH
Carnegie-Mellon University

APPENDIX

In their 1979 article, Hylland and Zeckhauser pursue a heuristic argument [due to Zeckhauser, "Group Decision and Allocation," Discussion Paper #51 (1968), Harvard Institute of Economic Research] to establish the impossibility of a rule for amalgamating a set of Bayesian preferences into a single Bayesian preference, provided that the rule: (1) applies to all potential sets of Bayesian agents; (2) respects the weak Pareto condition; (3) avoids "Dictators" for the group-probability; (4) combines probabilities and utilities separately; and (5) preserves unanimity of the agents' probabilities in case they all have the same degrees of belief. Also, they show that (2) and (3) may be replaced by the single condition (6): the strong Pareto requirement.

We note, first, that regarding "impossibility," the two results share an

[21] "A Representation for Preference as a Strict Partial Order in terms of Sets of Pairs of Probabilities and Utilities," *op. cit.*

[22] In his doctoral thesis, "Existence of Compromises in Simple Group Decisions" (Dept. of Statistics, Carnegie Mellon University), Goodman extends the negative finding of theorem 1 to groups of three agents ($n = 3$). Also, he gives sufficient conditions that only autocratic "compromises" exist with groups of n-many Bayesian agents.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

equivalence between the conjunction of conditions (2) weak Pareto and (3) no (probability) dictators/no autocrats, and condition (6) the strong Pareto requirement. (The weaker condition, "no (probability) dictators," suffices in their argument because they have condition (4). Instead, we use "no autocrats.") Unlike their result, however, theorem 1 does not place restrictions on the form of the amalgamation, i.e., neither condition (4) nor (5) is required for our analysis. Thus, it is insufficient for finding a Bayes compromise merely to abandon independent amalgamations of probability and utility.

Also, theorem 1 shows that there are none but autocratic, weak Pareto "compromises" for *every* pair of agents with differing beliefs and values. This is in marked contrast with their result which states that, for each Bayesian social welfare rule satisfying 3–5, there is some pair of Bayesian agents whose preferences are amalgamated in violation of (2): the weak Pareto condition.

A somewhat different treatment of the Bayesian group decision problem is found in Hammond's 1981 paper. Hammond's interesting work is concerned with issues of welfare optimality in dynamic (intertemporal) social decisions. Abstracting away from the dynamic features of his analysis, we find the following result about Bayesian amalgamations:

"*Theorem 2* " (Hammond, *op. cit.*, p. 241): Suppose there are n Bayesian agents, whose preferences over their own gains from social acts are represented by the probability/utility pairs (P_k, U_k), $k = 1, \ldots, n$. Then, a Bayesian (Bergson) social welfare function, W, satisfying the strong Pareto condition exists, provided it is of the form (P_W, U_W), where: (i) $P_W = P_k$ $(k = 1, \ldots, n)$, i.e., the n agents all have the same personal probabilities, and (ii) $U_W = \sum_k \gamma_k U_k$ $(k = 1, \ldots, n$ and $0 < \gamma_k)$, i.e., social utility is a convex combination of the n individual utilities. Thus, according to this result, there can be no Bayesian amalgamation (subject to the strong Pareto condition) whenever different agents hold different personal probabilities, regardless of the nature of their personal values for outcomes.

Theorem 1 is stronger than this result because, as in the previous comparison, Hammond's conclusion depends upon a restricted form of group amalgamation—a Bergson social welfare rule makes the group utility a function of the individual (interpersonal) utilities, independent of their personal probabilities. A Bergson amalgamation treats beliefs and values separately. (There is, in addition, an assumption that the amalgamation of individual utilities is differentiable, i.e., Hammond's argument requires a smoothly changing Bergson social welfare.) Also paralleling the comparison with Hylland and Zeckhauser's result, Hammond's argument shows that, for every Bergson social welfare rule (unless the agents share a common personal probability), there is some configuration of personal utilities which leads to a violation of the strong Pareto condition. Theorem 1, by contrast, establishes the violation of the strong Pareto condition for each Bayes model and for every pair of agents with different beliefs and values.

Finally, it is worth explaining away an apparent conflict between Hammond's theorem 2 and other findings we have made, concerning Bayesian compromises when values are shared (there is a common utility) and beliefs differ (cf. "Decisions without Ordering," §4). "Theorem 2" asserts that no Bergson social welfare rule is possible in this case. Our discussion indicates

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

that there exist weak Pareto compromises created by the following pairs: Let the Bayes model use the (assumed) common utility and any convex combination of the personal probability distributions. In the case of two agents, each nonextreme convex combination of the individual probabilities (no one individual is autocrat) creates a compromise that satisfies the strong Pareto condition. [This generalizes to n agents. For conditions under which the strong Pareto compromises require all positive weights, see, e.g., P. C. Fishburn, "On Harsanyi's Utilitarian Cardinal Welfare Theorem," *Theory and Decision*, XVII (1984): 21–28.]

The two claims are not contradictory: they deal with different domains of social acts. In Hammond's presentation, the class of social acts includes the (smaller) domain of our analysis. For Hammond's "Theorem 2," the domain of social acts arises by feasible market transactions. In the restricted domain of our analysis, each agent receives the same outcome as every other agent, in every state. Thus, when we suppose that agents share a common utility for individual rewards, perforce it is also a common utility over the "constant" social acts in this limited variety of social choices.

Instead, the social acts used in Hammond's analysis include the commonplace situation where different individuals receive different rewards. The personal utilities of Hammond's analysis reflect the agent's preferences for her individual rewards only. Thus, for Hammond's argument, two agents with the same individual utility will differ in their preferences for some "constant" social acts. If they receive different rewards, one from another, their preferences between two social acts may be in direct opposition despite the common utility for individual rewards. For example, they can have the same (cardinal) utility for money but differ in their ranking of two social options, depending upon which of them receives the greater monetary reward under which of the two acts.

We can extend theorem 1 to the larger domain of Hammond's social choices by recasting our notion of a social option. Let us mean, rather, that an act is an n-tuple of horse lotteries, one for each agent. This is inclusive of the restricted class of social acts in which each agent receives the same horse lottery. Then, as we suppose when arguing that a convex combination of personal probabilities creates a weak Pareto compromise, in order for two agents to share a common utility over outcomes—over the "constant" acts in this extended sense of "option"—they must agree on the ranking of all social acts which award distinct (von Neumann–Morgenstern) lotteries to different individuals. If they have the same utility for these "constant" social acts then [in J. C. Harsanyi's sense; cf. Rational Behavior and Bargaining Equilibrium in Games and Social Situations (New York: Cambridge, 1977), ch. 4] they share one "moral" utility over such options. In that case, an agent's preference over "constant" social acts does not depend upon his identity; it does not depend upon which (lottery) reward is his. But when agents have a common utility for personal rewards and evaluate social acts solely in terms of their own gains (as in Hammond's analysis), their preferences do *not* correspond to a "moral" utility. Thus, the appearance of a conflict between the two claims is illusory.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

SCIENCE AND OBJECTIVITY

A central goal in doing epistemology of science is the validation of scientific knowledge, or at least an appraisal of the possibilities and attempts to validate scientific knowledge. This facet of philosophy of science seeks to answer the straightforward question, "Why believe what science claims about the world?" This is a question made more acute with the realization that science is largely a self-regulating system. Insofar as observation is theory-relative in the sense that theory influences not only what observations are to be made but also what those observations mean, the accountability of scientific claims is an internal affair and the reliability of science is self-proclaimed. So why should we believe science?

I. INTRODUCTION: OBJECTIVITY AND INDEPENDENCE

John Ziman[1] articulates a common and reasonable answer in claiming, "[t]he primary foundation for belief in science is the widespread impression that it is *objective*" (*ibid.*, p. 107). Israel Scheffler makes a similar point in *Science and Subjectivity*,[2] indicating that science pursues an ideal of objectivity and is credible to the extent that it nears that ideal. But what exactly is objectivity and how do we know when it is present? Ziman links objectivity to intersubjectivity, that is, agreement among many persons, in this case scientists. Scientific knowledge is believed to be reliable, because it is required to pass judgment by many disinterested (that is, objective) reviewers. Scheffler associates objectivity with a more epistemic notion of independence. Evidence is objective if its gathering is independent of that for which it serves as evidence.

These are provocative suggestions, though they are still vague. My plan here is to pursue the notion of epistemic independence by making it more precise and showing how it can function in assessing our warrant for belief in scientific claims. If we think of objectivity as requiring an independent appraisal, then it is desirable to characterize precisely the relevant notion of independence and to indicate how it can be recognized.

This kind of epistemic independence can be applied in a variety of situations of validating scientific knowledge. One such situation is that of observations. Science is reliable, we might think, because its claims must be tested against the objective facts of observation. But by now many philosophers have conceded to a certain amount of theory dependence in observation and its role as an objective stan-

[1] *Reliable Knowledge* (New York: Cambridge, 1978).
[2] Indianapolis: Bobbs-Merrill, 1967.

0022-362X/89/8605/245–257 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

dard is threatened. Yet even if we cannot have observation that is entirely theory-independent, we might at least have observation that is independent of the particular theory we are testing. That is, if some observation O is to be used in confirming a theory T, then a degree of objectivity is preserved if the theories on which O depends are more or less independent of that particular theory T. Objective evidence is evidence that is verified independently of what it is evidence for. Ian Hacking,[3] in discussions of observation, refers to this phenomenon as, "the disunity of science" (*ibid.*, p. 183). Science is disunified in the sense that it is composed of several independent branches. It is a healthy disunity, since the independent branches can assist in independent, objective testing of each other's claims.

The benefits of independence can be appreciated by considering our own human perceptual systems. We consider our different senses to be independent to some degree when we use one of them to check another. If I am uncertain whether what I see is a hallucination or real fire, it is less convincing of a test simply to look again than it is to hold out my hand and feel the heat. The independent account is the more reliable, because it is less likely that a systematic error will infect both systems than that one system will be flawed.

The independence of sensory systems is a physical kind of independence, in the sense that events and conditions in one system have no causal influence on events and conditions in another. But the independence relevant to objectivity in science is an epistemic independence between theories. A closer analogy to the epistemic concept is found in history. Ancient history in particular is fortunate to have available two independent sources of information, the literature of ancient texts and archeological artifacts. Interpretation of artifacts, their age, their use in trade, and the like, is credible insofar as it is consistent with interpretations of ancient texts. In this way, claims made by archeologists can receive an independent corroboration by the claims made by interpreters of ancient texts. But the precise degree to which the two sources, archeology and literature, are independent is an important question, and it is just this issue, in the context of the sciences in general, which will be pursued here.

There are plenty of good examples in the physical sciences of independence functioning as a rational measure of belief in scientific claims. One of the best is in the experiments of Jean Perrin in the beginning of this century, experiments designed to measure the number of molecules in one mole of material, Avogadro's number.[4]

[3] *Representing and Intervening* (New York: Cambridge, 1983).

[4] A thorough account of Perrin's accomplishment is to be found in M. Nye, *Molecular Reality* (London: Macdonald, 1972).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Perrin measured the same physical quantity in a variety of different ways, thereby invoking a variety of different auxiliary theories. And the reason that Perrin's results are so believable, and that they provide good reason to believe in the actual existence of molecules, is that he used a variety of *independent* theories and techniques and got them to agree on the answer. The chances of these independent theories all independently manufacturing the same ficticious result is small enough to be rationally discounted. It is the independence that supports the credibility of the account. But what exactly does it mean to claim in this context that two theories are independent?

II. INDEPENDENCE OF THEORIES

The concept of independence of two theories which is relevant to objectivity must be applicable to epistemic justification. Circumstances of discovery of the theories, for example that they were first proposed by the same person or motivated by the same evidence, should not influence the evaluation of independence in this sense. If an Albert Einstein is clever enough to propose both the special theory of relativity and the quantum theory of light, this does not make the theories any less epistemically independent. One could still supply an objective, independent check of the other as long as the truth conditions of one do not influence the truth conditions of the other.

Given two scientific theories T_1 and T_2, T_1 is independent of T_2 in a way that makes T_1 a possible source of objective test of T_2, if our acceptance of T_1 as true (or rejection of T_1 as false) does not force us to accept T_2 as true (nor to reject T_2 as false). If the truth or falsity of T_1 is insulated from the truth or falsity of T_2, then the two theories are independent in the relevant way. T_1 could function as an auxiliary theory in an observation in support of T_2 without our worrying that T_1 or T_2 is acting in its own self-interest. This is what is meant by objective testing.

The relevant notion of independence then is one in which the acceptability of T_2 as true (or rejection as false) is not a function of the acceptability of T_1 as true (or rejection as false). That is, the acceptance as true or rejection as false of any one element of T_1 has no repercussion for the truth or falsity of T_2. To clarify this general idea of independence and to be able to recognize independence when it is present, we need a more careful specification of what scientific theories are, what it means to say that one is accepted as true or rejected as false, and how the truth or falsity of one theory can affect that of another.

Consider a theory to be described by a set of propositions. This is not to say that a theory simply *is* a collection of propositions, but if a

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

theory is to be an issue of discussion and an object of belief, it must make propositional claims. Whatever one's views on the nature of scientific theories, whether they should be construed syntactically or semantically, a theory must be presented as a body of sentences, and it is this presentation that will be the object in describing epistemic independence.[5] Thus, from now on, reference to a theory T will be intended as reference to the presentation of T, the set of propositions which describe the theory.[6]

To say that a theory is accepted as true will mean that one accepts as true all of the propositions in the description of the theory. To say that a theory is rejected as false, on the other hand, will mean that at least one of the propositions is rejected as false. It will be useful to employ the notion of a *consistent negation* of a theory. If a theory is described by a set of propositions $\{P\}$, then a consistent negation of T $(-T)$ is the set of propositions with some or all of the Ps negated, with the restriction that the resulting set is consistent.[7] So T is rejected as false if one endorses (accepts as true) a consistent negation of T.

Attending to the sets of propositions which describe theories allows for a precise characterization of independence of theories. T_1 is independent of T_2, if T_1 being true (or false in any way) does not force us to accept T_2 as true (or reject T_2 as false). That is, T_1 does not entail either T_2 or $-T_2$, nor does $-T_1$ entail either T_2 or $-T_2$. Furthermore, if the acceptability or rejection of T_2 is to be entirely isolated from T_1, we cannot allow T_1 or $-T_1$ to function as essential assistance in proving (or disproving) parts of T_2 from other parts of T_2.[8] So, if I accept the theory T_1, as well as some parts of T_2, and this allows me to prove other parts of T_2, then the two theories cannot be wholly independent.

These constraints on independence can be marshalled into a definition of theory independence. This definition is an adaptation of a

[5] The importance of the role of the presentation of a theory is described and defended in James Griesemer, "Presentations and the Status of Theories," in P. Asquith and P. Kitcher, eds., *Philosophy of Science Association 1984*, vol. 1 (East Lansing, MI: Philosophy of Science Association, 1984).

[6] It is likely that a full explication of what a theory *is* will include further requirements, such as deductive closure. But an exact spelling out of these requirements is unnecessary here, since our notion of independence will operate on any consistent collection of propositions and will accommodate subsequently placed restrictions on the collection.

[7] As an example of a consistent negation, consider the theory $T = \{p \supset q, q \vee r, r\}$. Then one possible consistent negation of T is the set $-T = \{p \supset q, q \vee r, -r\}$, but the set $\{p \supset q, -(q \vee r), r\}$ is not a possible consistent negation, since it is not consistent.

[8] This possible manifestation of theoretical dependence I realized in reading Nicholas Rescher, *The Coherence Theory of Truth* (New York: Oxford, 1973), p. 335.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

notion of independence between sets of propositions presented by Nicholas Rescher.[9] Given two theories T_1 and T_2, described by sets of propositions $\{P\}$ and $\{Q\}$ respectively, say that T_1 is independent of T_2 if and only if there is no P in T_1 and no Q in T_2 such that there is a consistent set T with either $T \subseteq (T_1 \cup T_2)$ or $T \subseteq (-T_1 \cup T_2)$ and such that $(P \in T \vee -P \in T)$, and $(T \vdash Q \vee T \vdash -Q)$, and finally there is no T' such that $T' \subset T$ and $(T' \vdash Q \vee T' \vdash -Q)$. Thus, T_1 is independent of T_2 just in case there is no subset of either T_1 or a consistent negation of T_1 which entails an element of T_2 or the negation of that element, nor can the subset be used with a subset of T_2 to entail another element of T_2 or its negation. T_1 being true or false in any way does not force (entail) the truth or falsity of T_2.

It is not necessary to follow all the details of the above definition of theory independence to appreciate the concept and its role in objectivity. But realizing that the intuitive notion of independence as the truth assessment of one theory being isolated from the truth assessment of another is spelled out in terms of sets of propositions and entailment is enough to indicate some interesting features of this characterization. There is some element of caprice, for example, in determining which propositions are to be included in the collection that describes the theory. If there is some proposition Q in the description of T_2 such that $-T_1 \vdash -Q$, then one might simply jettison Q from the description of the theory and deal with a new, edited version of T_2 which is independent of T_1.[10]

As a simple example, let $T_1 = \{t \supset p, r \supset p, s\}$ and $T_2 = \{p \supset q, p \vee s, q\}$. Then T_1 and T_2 are not independent, since one possible consistent negation $-T_1$ is $\{t \supset p, r \supset p, -s\}$ and, by forming $T \subseteq (-T_1 \cup T_2)$ where $T = \{-s, p \vee s, p \supset q\}$, it is the case that $T \vdash q$. But T_1 could be edited to $T_1^- = \{t \supset p, r \supset p\}$ and then it is the case that T_1^- is independent of T_2. The description of independence of theories, as it has been presented so far, is really a description of independence between consistent sets of propositions in general. It will have to be combined with some standard of determining which propositions must be included to form a whole theory to avoid sacrificing elements for the sake of independence.

But there is a second notion of independence which will make use of the independence of theories in a way which precludes this question of which propositions must be included and which is more directly applicable to actual cases of objective testing. I shall refer to this new idea as *independence of an account*, since it will be used to evaluate the epistemic independence of the theory or theories used to account for (to validate) observational data as evidence. This will

[9] *Op. cit.*
[10] This caprice will be somewhat limited under a constraint of deductive closure.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

respond to the question of what it is to have independent evidence for a theory.

III. INDEPENDENCE OF AN ACCOUNT

To claim generally that observation is theory–dependent hides a great deal that is relevant to issues of epistemology and science. If observation is theory–dependent, on which theories does it depend and in what ways? The responses to these questions are necessarily case–by–case affairs. Different observations may depend on different theories and in ways that differ in epistemologically significant ways. It is only by describing the nature of specific cases of observation that one can appreciate the impact of the theory dependence of data.

Several such studies of specific cases of observation have been done. Dudley Shapere[11] describes observations of the sun which depend heavily on the observer's theoretical background, and Hacking[12] explains how theory accounts for the information scientists claim to be getting through optical microscopes. With greater emphasis on human perception and less on machine-aided observation, Jerry Fodor[13] suggests an empirical approach to determining at least which theories in our conceptual background do *not* influence particular observations.

In all of these examples, an observation will be theory-dependent in the sense that there is some collection of theories which account for the information one claims to receive in the observation, the information that functions as evidence in science. These theories influence the observation in that a change in the theories would force a change in the claim about the observational information. Thus, for any singular claim that is based on an act of theory-dependent observation, there will be a collection of theories which validate the claim. Call this the set $\{T\}$ of theories which account for the particular observation.

Sometimes the set $\{T\}$ is easy to identify. The examples by Shapere and Hacking demonstrate this feature. Take as our own example the observation with an electron microscope of a DNA molecule, which we will follow through the description of independence of an account. The accounting theories in this case are those which describe the interactive properties of electrons as they scatter off the specimen and are focused in the microscope and impart their energy to the viewing screen. This is a substantial set of sophisticated theories and it represents a major capitulation to the theory dependence of

[11] "The Concept of Observation in Science and Philosophy," *Philosophy of Science*, XLIX (1982): 485–525.
[12] *Representing and Intervening* (New York: Cambridge, 1983), ch. 11.
[13] "Observation Reconsidered," *Philosophy of Science*, LI (1984): 23–43.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

observation. But it need not compromise the objectivity of the evidence or the objective nature of the test it provides. The test, after all, is to be of some theory about the specimen, a theory of DNA, and a degree of independence in the account, that is, objectivity of the evidence, is preserved if the singular observational claim is supported by an account that is independent of the theory being tested.

In characterizing this kind of independence, independence of an account, it will be necessary to refer to the theory T_x that is being tested. The idea is that, if one is making observations of xs (for example, DNA molecules), then these observations will function as evidence in confirming the theory-of-x, T_x. It will also be useful to identify T_x1 as that specific subtheory (subset of propositions) of T_x for which the singular observation claim could be used as confirmation. That is, T_x1 will be that collection of propositions taken from T_x for which the fact disclosed by the observation in question could function as confirming evidence. The identification of T_x1 will depend on a conception of confirmation which can provide a decision procedure for determining whether or not some particular evidence confirms a particular hypothesis (described by a proposition). For this reason, the membership of T_x1 will be relative to one's operating conception of confirmation. Furthermore, the concept of independence which results will not then be useful in formulating the specifics of confirmation. We might have hoped to use the complement of independence to specify a kind of relevance between theory and evidence appropriate to confirmation. But this application of the concept of independence has been blocked.

As an example of a theory T_x and the identification of the subcollection T_x1, consider that branch of the scientific description of light referred to as geometric or ray optics. This is T_x, where the xs are light rays. T_x will include such statements as Snell's law of refraction, a description of total internal reflection, the law of reflection, and the like. If one observes an image in a mirror and locates the image at the position predicted by ray-optical calculations, then the observation functions as a hypothetical-deductive confirmation. So, relative to a hypothetico-deductive conception of confirmation, and with the particular observation being of mirror imagery, the law of reflection is a member of the subtheory T_x1, since it is an essential part of the deductive prediction that proved to be true. The observation of mirrors can hypothetically-deductively confirm the law of reflection but not Snell's law, since the latter, in its description of refraction, has no impact on the deduction in question. Snell's law is not a member of T_x1. The membership of T_x1 will be a function of one's ideas about confirmation and of the particular observation of which the independence and objectivity are a question. To insure indepen-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

dence, it is the particular subtheory T_x1 that one wants to isolate from the theoretical account of the observation.

The next step in the characterization of independence of an account is to identify another subtheory T_x2 of T_x as that subset of propositions in T_x which is not independent of the collection of theories $\{T\}$ used to validate the singular observation claim. This makes use of the concept of independence between theories as described in the previous section, since we will test to see if the theory T_x2 is independent of the theory that is the conjunction of the theories in $\{T\}$. But there is no capriciousness of inclusion in these theories, since $\{T\}$ is specified by what is required to account for the observation in question, and T_x2 is by design to include any part of T_x which fails the test of independence with $\{T\}$.

The motivation for identifying the subtheories T_x1 and T_x2 is the realization that for some theory-of-x T_x there may be parts T_x2 which are used in the account of evidence which functions in confirmation of other parts T_x1. The concern in evaluating the independence of an account is in isolating the potential beneficiary of the evidence, T_x1, from the guarantor of validity of the observation, $\{T\}$. The potential breach in that isolation is through the home theory's participation in the account, the T_x2. This evaluation of independence of the account reduces to an assessment of the relationship between T_x1 and T_x2 and of the relationship between T_x2 and $\{T\}$.

Examples of this kind of division of a theory into two parts are commonplace. The case of observing the DNA molecule is such an example. If the putative observation is of the shape of the molecule, then T_x1 is that part of the theory of DNA which describes molecular shape. And if the observation is by way of electron microscope, then we must account for the molecule as an appropriately massive scattering target for electrons. Hence, T_x2 will be that part of the theory of DNA which describes molecular composition.

Identifying the subtheories T_x1 and T_x2 will always require a detailed analysis of the actual observation claim to see just what claim is being made (to identify T_x1) and how the observation is done and how it is validated (to identify T_x2). For this reason, it is difficult to answer a general question of the objectivity of the testing of a theory T_x, let alone of the objectivity in science in general. But if we focus on a particular observation, say, observation of the motion of tectonic plates, then the independence of the account as an indication of objectivity of the evidence is a straightforward calculation. It will procede by identifying T_x1 as that part of tectonic theory which describes and predicts plate motion, and T_x2 as that part of the theory which describes the composition of the plates and their char-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

acteristics of interaction with the seismic waves that are used as probes.

The point in this description of the concept of independence is twofold. For one, the assessment of independence of an account can only be done by describing the details of the particular case of observation. One cannot make general claims as to objectivity, insofar as it is based on this measure of independence, without the support of evidence from a variety of specific case studies. Second, whole theories are not the appropriate units of appraisal of independence of an account. The concept of independence which is relevant to evaluating the objectivity of evidence requires that we group some theories, the $\{T\}$, and split up another, T_x. It is a likely fact of life that some part of the theory-of-x will be used to account for the observation of some aspect of a particular x. But a degree of independence is preserved if this is not a crucial part of the theory, the part which explicitly discusses that aspect of x which has putatively been observed, the part which could therefore benefit (by confirmation) from the observation.

In relying on the relationship between T_x1 and T_x2 and the relationship between T_x2 and $\{T\}$, this characterization of independence of an account potentially admits of degrees. The clear demarcation of independence as alluded to above is simply that T_x1 and T_x2 are disjoint. One part of a theory supports the evidence used to confirm *another* part of the theory. But on either side of this demarcation, toward more and toward less independence, we can identify varying degrees of the condition of independence. For example, T_x2 could have varying degrees of involvement with the set $\{T\}$ of accounting theories. Since T_x2 is that subcollection of propositions which are not independent of the $\{T\}$, it is possible that T_x2 is not even a member of the $\{T\}$ but is merely not independent. More likely though is that T_x2 is one of many members of $\{T\}$, that is, that the set T_x2 is a proper subset of $\{T\}$. And it is not impossible that T_x2 is all that is required to account for the observation and no outside help is enlisted. In this case, T_x2 is identical to the set $\{T\}$ and one is inclined to regard the resulting lack of independence as more serious than if T_x2 is forced to cooperate with other branches of science.

One could also find a scale of independence in the varying degrees of closeness of relation between T_x1 and T_x2. It would be ideal, from the standpoint of independence, if T_x2 were empty and no part of the theory-of-x participated in the observation. But much more likely is that T_x2 is a nonempty, disjoint crony of T_x1. This is the case we have been dealing with, in which a theoretical crony participates in the evidential account. Beyond these limits and into the realm of

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

nonindependence, that is, if T_x1 and T_x2 are not disjoint, we might regard the case as worse (less independent) if T_x1 is actually a subset of T_x2 than if the two sets merely intersect. The case of subset would indicate that the account of observation of the particular x is such that no theoretical claim about the x and the property observed is independent of the theories used to justify the observation. The observation claim is, by standards of independence, indistinguishable from a theoretical claim. But the worst case of all would be if T_x1 were identical to T_x2 and T_x2 were in fact the whole of $\{T\}$. This case is utterly without independence and is blatantly devoid of objectivity in the evidence for the theory T_x.

The suggestion is that the closeness of relation between T_x1 and T_x2 as well as the depth of involvement of T_x2 in $\{T\}$ could be used to provide a scale of independence of account. But it is better not to get involved here in the details of this metric of independence. A possible thermometer of independence of account is presented in the appendix which follows, but it will be more productive now to show some examples of the observation claims and their evaluation in terms of independence of account.

IV. EXAMPLES

In developing a concept such as independence of an account and its relevance to objectivity, one wants a characterization that agrees with intuitive assessment of easy examples so that we can use it with confidence as a guide to intuition in the difficult cases. That is, one wants a characterization of independence which gets a high score, indicating an objective test of theory, in cases that we are predisposed to believe as cases of objective testing and reliable information. And where we have antecedent reason to disbelieve, cases of unreliable information, the measure of independence ought to get a low score. Only then will the notion of independence of an account be acceptable as a warrant for belief in cases where we have no other indication of reliability of information.

With this in mind I present three examples of the concept of independence of an account in action, two in which we have reasons for prejudgment and one in which we do not.

The best case, in the sense of the most independence of an account, is the case where T_x2 is empty. No part of the theory-of-x is used in the account. Observation claims of what we unhesitatingly call observables are often of this degree of independence. Consider the observation claim that the stone is accelerating. The theories from which one would draw to corroborate this claim, that is, the scientific account of the reliability of the observation, would be theories of optics and perception. But no part of either geology or, more importantly, kinematics, appear in the account. This is just to

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

say that T_x2 is empty. Thus, in this and other cases of perception of midsized objects, our pre-analytic intuition accepts the observation as reliable information and the analysis of independence concurs with a ruling of high independence indicating a relatively objective test.

The worst case, that is, the least independent account, results if T_x1 is identical to T_x2, which in turn is the whole of $\{T\}$, that is, if $\{T_x1\} = \{T_x2\} = \{T\}$. This means that all and only the theory-of-x and the particular property putatively observed is used to account for the observational reception of the information about a particular x having this property. The second link in the equation, the 'only' part of 'all and only', is unlikely to occur in actual science. It is rare that no disinterested auxiliary theory, usually a theory of optics or perception, is used in accounting for the evidence. But as a case that comes close to the bottom of the scale of independence, consider a historical account of alleged observation of caloric fluids. More precisely, we should describe this as a claim of the acquisition of information about caloric.

Use as an observing apparatus some solid object such as a block of metal. The particles of caloric which, according to the caloric theory, are present in the object will be distributed throughout the object as a result of the repulsive force between particles. The caloric particles will surround the particles of the object.[14] As more caloric is added to the object, the ether-cloaked object particles experience a repulsive force between themselves as the ether particles push for distance between each other. The more caloric, the greater repulsive force and the object responds by expanding.

In this way, the expanding block displays information about caloric fluid, namely, that caloric is flowing into the block. But the informational link between the appearance of a block of metal and properties of caloric is provided almost entirely by the caloric theory itself. In other words, the set $\{T\}$ of accounting theories is constituted almost entirely of claims T_x2 drawn from ether theory. And any property of ether, its characteristics of flow or its effects on material particles, must be described in the theoretical account in order to support the observational account. To adapt a Wittgensteinian analogy, we are not even buying another copy of the newspaper to check its reliability; we are simply accepting the assuring words printed at the top of each page, "all this is true."

The moral of the story is that, even though the ether theorist can claim observational evidence for the ether theory, that evidence can be discounted on the basis of its lacking independence. It is not

[14] This piece of caloric lore is adapted from a description in G. Cantor and M. Hodge, eds., *Conceptions of Ether* (New York: Cambridge, 1981), pp. 27/8.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

objective evidence. In this case, again, our intuition (that the ether evidence is unreliable) parallels the analysis of independence of the account.

The majority of evidential claims in science fall somewhere in between the best and worst and are not immediately decidable with respect to reliability. In these cases the independence of account is useful as a measure of warrant for belief. To demonstrate this we can resurrect the DNA molecule as imaged with an electron microscope which was considered above. The image by the electron microscope may reveal, for example, that the DNA is denatured (split). In this case, T_x1 is that part of the theory of DNA which discusses the causes and effects of the splitting of the molecule. T_x2 will not be empty, since there must be an account of the composition of the molecule to explain its differential absorption of stain and its properties of scattering electrons. But T_x2 in this case appears to be disjoint from T_x1. And, of course, T_x2 is not the only source of theoretical support for the observation, since a great deal of electron scattering theory and electromagnetic theory are invoked to trace the production of the image.

This demonstrates that subdividing the theory of DNA into the claims about shape of the molecule (T_x1) and claims about composition of the molecule (T_x2) reveals a degree of independence in the account of electron-microscopic observation of a DNA molecule. The image on the screen of the microscope can therefore be regarded as objective evidence for the theory of DNA. This assessment, though, is really only a first-order approximation. The fine structure of independence in this account could be described through a characterization of the varying degrees of independence as suggested in the appendix.

V. CONCLUSION

The point of this discussion has been to suggest a characterization of objectivity. It has been motivated by the claim that objectivity is what makes science believable and is in part what is scientific about science. An objective test, we assume, is one that is less likely to reproduce theoretical artifacts and propagate hidden systematic error. An objective test, that is, is a more reliable test than is one administered by the testee. External review is valued for its trustworthiness. And objectivity in this sense can be evaluated in terms of accessible features of a scientific report, namely, independence of an account. There is no reference to external, inaccessible features such as truth or facts of the matter, and, for this reason, independence, in its relevance to objectivity, is a valuable and workable standard for evaluating warrant for belief.

The relevant concept for this purpose is independence of an ac-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

count. The application of this concept is specific to each case of putatively objective evidence and requires dividing the theory to be tested into a part that will directly benefit from the test and a part that participates in the testing. This approach then maps neatly on to many examples from actual science.

The discussion has been done with the realization that observation in science is informative only with the aid of theory. In terms of epistemic compromise and of threatening reliability, though, some theory dependences are worse than others. A theory/observation dichotomy (or an observable/unobservable dichotomy) is therefore of limited importance as a guide to belief, both because so many observations are so intimately theoretical and because it hides the detail (namely, independence) that is significant in evaluating warrant for belief. My suggestion, then, is to replace the traditional theory/observation dichotomy with an independent/not-independent dichotomy (or spectrum) as an indication of what is epistemologically significant in scientific evidence.

PETER KOSSO

Northwestern University

APPENDIX

A way of scaling the degrees of independence of an account.

Closeness of relation between T_x1 and T_x2	Depth of involvement of T_x2 in $\{T_i\}$
$T_x2 = \phi$ most independent	
$(T_x2 \cap T_x1) = \phi$ theoretical cronies	$T_x2 \notin \{T_i\}$ $T_x2 \in \{T_i\}$ but not $\{T_x2\} = \{T_i\}$ $\{T_x2\} = \{T_i\}$
$(T_x2 \cap T_x1) \neq \phi$ but not $T_x1 \subseteq T_x2$ theoretical nepotism	$T_x2 \notin \{T_i\}$ $T_x2 \in \{T_i\}$ but not $\{T_x2\} = \{T_i\}$ $\{T_x2\} = \{T_i\}$
$T_x1 \subseteq T_x2$ but not $T_x1 = T_x2$	$T_x2 \notin \{T_i\}$ $T_x2 \in \{T_i\}$ but not $\{T_x2\} = \{T_i\}$ $\{T_x2\} = \{T_i\}$
$T_x1 = T_x2$ least independent	$T_x2 \notin \{T_i\}$ $T_x2 \in \{T_i\}$ but not $\{T_x2\} = \{T_i\}$ $\{T_x2\} = \{T_i\}$

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

BRUTE EXPERIENCE*

THE question whether brutes have experiences has been granted as obvious in recent times, and in one sense of the term 'experience' no doubt it is so. But not, I shall argue, in the sense that makes their experiences an appropriate object of moral concern.

I.

Since Thomas Nagel's seminal paper "What is it Like to be a Bat?",[1] it has become generally accepted that a creature may be said to have experiences if and only if there is something that it is like to be that thing (even if we cannot know what). But this identification of experience with subjective feel is false. There are, in fact, many experiences that do not feel like anything.

Consider some familiar examples. While driving the car over a route I know well, my conscious attention may be wholly abstracted from my surroundings. I may be thinking deeply about a current piece of writing of mine, or phantasizing about my next summer's holiday, to the extent of being unaware of what I am doing on the road. It is common in such cases that one may suddenly "come to," returning one's attention to the task at hand with a startled realization that one has not the faintest idea what one has been doing or seeing for some minutes past. Yet there is a clear sense in which I must have been seeing, or I should have crashed the car. My passenger sitting next to me may correctly report that I had seen the lorry double parked by the side of the road, since I had deftly steered the car around it. But I was not aware of seeing that lorry, either at the time or later in memory.

Another example: when washing up dishes I generally put on music to help pass the time. If it is a piece that I love particularly well, I may become totally absorbed, ceasing to be conscious of what I am doing at the sink. Yet someone observing me position a glass neatly on the rack to dry between two coffee mugs would correctly say that I must have seen that those mugs were already there, or I should not have placed the glass where I did. Yet I was not aware of seeing those mugs, or of placing the glass between them. At the time I was swept up in the finale of Schubert's "Arpeggione Sonata," and if asked

* I am grateful to Clare McCready for discussions that stimulated the writing of this paper, and to A. D. Smith for his comments on an earlier draft.

[1] *Philosophical Review*, LXXXIII, 4 (1974): 435–451.

0022-362X/89/8605/258–269 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

even a moment later I should have been unable to recall at what I had been looking.

Let us call such experiences *nonconscious experiences.* What does it feel like to be the subject of a nonconscious experience? It feels like nothing. It does not *feel like anything* to have a nonconscious visual experience, as of a lorry parked at the side of the road or as of two coffee mugs placed on a draining rack, precisely because to have such an experience is not to be conscious of it. Only conscious experiences have a distinctive phenomenology, a distinctive feel. Nonconscious experiences are those which may help to control behavior without being felt by the conscious subject.

These points—intuitive as they are—are already sufficient to show that Nagel is wrong to identify the question whether a creature has experiences with the question whether there is something that it feels like to be that thing. For there is a class—perhaps a large class—of nonconscious experiences that have no phenomenology. So, the fact that a creature has sense organs, and can be observed to display sensitivity in its behavior to the salient features of its surrounding environment, is insufficient to establish that it feels like anything to be that thing. It may be that the experiences of brutes (that is, of some or all nonhuman animals) are wholly of the nonconscious variety. It is an open question whether there is anything that it feels like to be a bat or a dog or a monkey. If consciousness is like the turning on of a light, then it may be that their lives are nothing but darkness. In order to make progress with this issue, we need to understand the nature of the distinction between conscious and nonconscious mental states.

Before proceeding to that task, however, it is worth noticing a somewhat less familiar example of nonconscious experience, since this will help us to see how the conscious/nonconscious distinction may have a physical realization in the neurological structure of the human brain. The phenomenon I have in mind is that of blindsight.[2] Human subjects who have suffered lesions in the striate cortex (the visual center in the higher part of the brain) may lose all conscious experience in an area of their visual field. They insist that they can see nothing at all within that region. Nevertheless, if asked to guess, they prove remarkably good at describing features of objects presented to them in that area, such as the orientation of a line, or at pointing out the direction of a light source. They can also reach out

[2] This has been researched by a number of investigators. See in particular L. Weiskrantz, "Varieties of Residual Experience," *Quarterly Journal of Experimental Psychology,* XXXIII, 3 (1980): 365–386.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

and grasp objects. Indeed, if asked to try to catch a ball thrown toward them from their blind side, they often prove successful.[3]

The conclusion to be drawn from these studies is that, while blindsight patients lack conscious visual experience within an area of their visual field, they nevertheless have nonconscious experiences which are somehow made available to help in the control of their actions. It seems that the neurological explanation for the phenomenon is that information from the eye is not only mapped on to the striate cortex (in normal subjects) but is also sent to a second mapping in the midbrain. It is presumably this latter mapping which is made available, in blindsight patients, to be integrated with the subject's goals and other perceptions in controlling behavior. It is also possible that it is this midbrain information which underlies the everyday examples of nonconscious experience outlined above. But we should beware of concluding that any creature with a striate cortex will be the subject of conscious visual experiences. The phenomenon of blindsight shows only that a functioning striate cortex is a physically necessary condition for conscious visual experience, not that it is sufficient. It may be that in the case of everyday nonconscious experience the striate cortex is indeed active, but that its information is not made available to whatever structures in the human brain underlie consciousness. And it may be that nonhuman animals with a striate cortex do not possess those structures at all.

It is worth stressing that the various nonconscious experiences we have considered do genuinely deserve to be counted as a species of *experience.* For not only is incoming information processed to quite a high degree of sophistication, but the states in question conform to the practical-reasoning model of explanation. Thus, the car driver behaved as he did because he *wanted* to reach his destination safely and *saw* that the lorry was an obstacle in his path. And the blindsight patient picked up the ball because he *wanted* to comply with the request of the experimenter and *saw* that the ball was on the edge of the desk. But if someone really insists that experiences are conscious states *by definition,* then the conclusion of this section may simply be rephrased. It is that, since there exist in humans similar levels of cognitive processing and behavior control to those displayed by

[3] For these last two facts, I rely upon personal communication from A. J. Marcel, Applied Psychology Unit, Cambridge, who will be publishing details shortly. Subjects displayed 80%–90% of normal accuracy in reaching out for objects of varying sizes and orientations, placed at varying distances. For examples of other experimental data suggesting the existence of nonconscious perceptions of quite a sophisticated sort, see Marcel, "Conscious and Unconscious Perception," *Cognitive Psychology,* XV, 2 (1983): 197–300.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

brutes, which do not involve experiences, it is an open question whether brutes have experiences at all. In the discussion that follows, however, I shall assume, as seems most natural, that not all experiences are conscious ones.

II.

What distinguishes conscious from nonconscious experiences? The question is best raised in connection with the distinction between conscious and nonconscious mental states generally. Since David Armstrong's[4] early work, it has been usual to characterize conscious mental states as those which give rise (noninferentially) to an activated second-order belief in their own existence. Thus, a conscious belief that *P* is one which, besides being available to enter into the causation of the subject's behavior, is apt to cause in them the activated belief that they believe that *P*.[5] Similarly, a conscious visual experience is one that, besides causing beliefs about the matter to which the experience relates, and being made available to nonconscious motor control processes, is apt to give rise to the belief that just such an experience is taking place.

If such an account were correct, then it would be very doubtful whether many species of animal could be said to enjoy conscious experiences. For only the most anthropomorphic of us is prepared to ascribe second-order beliefs to toads and mice; and many of us would have serious doubts about ascribing such states even to higher mammals such as chimpanzees.[6] At any rate, behavioral evidence for the possession of such states in higher mammals is contentious, whereas their absence from lower mammals, birds, reptiles, and fish is surely uncontentious. I shall show, however, that the proposed account is definitely incorrect. But this result is not a defense of conscious experience for brutes. Quite the contrary: the account of consciousness which emerges will make it even less likely that any nonhuman animals have conscious experiences.

I begin with an example I owe to Tim Williamson, designed to show that one cannot equate conscious believing that *P* with an activated second-order belief that one believes that *P*. In the course of a discussion of the merits and demerits of functionalism in the

[4] *A Materialist Theory of the Mind* (London: Routledge, 1968). Similar accounts have been defended recently in Peter Smith and O. R. Jones, *The Philosophy of Mind* (New York: Cambridge, 1986), and in my own *Introducing Persons* (Albany: SUNY, 1986).

[5] A belief is activated when it becomes engaged in the subject's current mental processes.

[6] For a useful review of the evidence of various levels of cognitive organization in a range animal species, see Stephen Walker *Animal Thought* (New York: Routledge & Kegan Paul, 1983).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

philosophy of mind, I might realize that I had for some time been speaking of functionalists as "we," also becoming angry when the views of functionalists were maligned, thus manifesting the activated second-order belief that I believe myself to believe in functionalism. But this might strike me with the force of self–discovery. If anyone had asked me previously whether I were a functionalist, I might have expressed uncertainty. In which case it would seem that the possession of activated second-order beliefs is not sufficient for conscious believing.

Another argument with the same conclusion is that the proposed account gets the focus of attention of conscious believing quite wrong. Conscious belief is surely world-directed in precisely the way that belief itself is. If I entertain the conscious belief that the world is getting warmer, then the primary object of my belief is the earth and its likely future temperature. Whereas if the proposed account were correct, the primary object of the conscious belief would be myself (I should be believing of myself that I possess a particular first–order belief), only focusing on the world indirectly, via the content of the first-order belief in question.

This point holds also for the proposed account of the distinction between conscious and nonconscious experience. Conscious visual experiences, too, are primarily world-directed. When I consciously see that there is a dagger on the desk before me, the primary focus of my attention is the dagger itself. In normal cases of conscious perception, our experiences are, as it were, transparent: representing the world to us without themselves being objects of attention. It is of course possible to pay attention to one's conscious experiences, as when I attempt a phenomenological description of my visual field. But this is a sophisticated and relatively unusual thing to do. Whereas on the proposed account it is the normal case: to perceive consciously that there is a dagger on the desk would be to have activated the belief about myself that I have an experience of there being a dagger on the desk.

Is it possible to do better? Indeed it is.[7] A conscious, as opposed to a nonconscious, mental state is one that is available to conscious thought—where a conscious act of thinking is itself an event that is available to be thought about in turn. (When we think things con-

[7] The account that follows is modelled on some of Daniel Dennett's suggestions in "Toward a Cognitive Theory of Consciousness," in *Brainstorms* (Montgomery, VT: Bradford, 1978), 149–173. But Dennett, implausibly, connects consciousness directly with the ability to give verbal reports. If there is such a connection, it needs to be established at least in part by empirical investigation, not fixed by definition.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

sciously to ourselves, the events that express our thoughts are themselves available to be objects of further thoughts—I can think to myself that my thought was poorly formulated, or hasty, or confused, or whatever.) In the case of belief, a conscious belief (qua standing state) is one that is apt to emerge in a conscious thinking with the same content. This is then able to handle Williamson's example: the reason why I had not consciously believed functionalism to be true is that I failed to have any disposition to think to myself, "Functionalism is true." The account also has the advantage that conscious beliefs have the same primary world–directedness as beliefs. For the conscious act of thinking, aptness to emerge in which is the distinctive mark of a conscious belief, is an event with the very same (world-directed) content as that belief itself. What makes my belief that the earth is getting warmer a conscious one is that I am disposed in suitable circumstances to think to myself, "The earth is getting warmer"; in both cases, the direction of focus is on the world, rather than on myself.

In the case of experience, a conscious experience is a state whose content is available to be consciously thought about (that is, which is available for description in acts of thinking which are themselves made available to further acts of thinking). In this case, I say 'available to be thought *about*', rather than 'apt to emerge in thinkings with the same content', because it is plausible to claim that most experiences have a degree of complexity and richness which may outreach our powers of accurate description. Nevertheless, every aspect of the perceived scene is made available to thought, even if only the thought that things are now subtly different. (Although the manner in which the leaves of a tree are shimmering in the breeze may defy description, I must at least be able to think to myself that the pattern of movement is now slightly altered, if it is.) Here, too, we can retain the primary world–directedness of conscious experience, since the normal way for information that is made available to thought through perception to emerge in acts of thinking is in thoughts about the object perceived—as when I think to myself that the dagger on the desk is richly ornamented.

When we turn to consider, not conscious experience of something in the world, but the more sophisticated state of consciousness of that experience itself, it is important to note that the suggested account is consistent with the existence of unanalyzable qualia. It may indeed be the case that the distinctive feel of my experience of a warm shade of red is incapable of further analysis, or even of nonrelational description. But I claim that what constitutes that feeling as a

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

conscious rather than a nonconscious state is that it is available to be consciously thought about. It is the status of qualia as conscious states, not the individual qualia themselves, which is being analyzed on the proposed account.

Besides the virtues mentioned above, my account provides a natural treatment of the examples of nonconscious experience with which we began. The reason why my perception of the double-parked lorry was not conscious, is that, while information about the lorry was somehow made available for integration into my actions, it was not available to my conscious thoughts. Similarly in the example of nonconscious perception of mugs on a draining board, what makes the experience nonconscious is that there was, in the circumstances, nothing available for me to think spontaneously about those mugs.

The issue of spontaneity is important in handling the blindsight examples. For although in these cases the visual information is, in a sense, available to be thought about (since if asked to guess what is there, subjects will generally guess correctly), it is not apt to give rise to spontaneous thoughts in the way that conscious experiences do. In the normal course of events, blindsighted people will have no thoughts whatever about objects positioned in the blind portion of their visual field. Indeed, when they do think about the matter, they are strongly inclined to believe that they see nothing.

One final virtue of my account is that it is able to explain why so many philosophers have been inclined to connect possession of conscious mental states with the ability to speak a natural language. For such a connection is at its most plausible (though still denied by many) where conscious thinkings are concerned. The idea that the ability to think things consciously to oneself is tied to the possession of a natural language has an immediate (if defeasible) plausibility. Whereas a similar thesis applied to the capacity for conscious experience seems much more puzzling. For why should it be supposed that language mastery is a necessary condition for a creature to enjoy conscious visual experiences? If the account sketched above is correct, then there may indeed be such a connection, but at one remove: it is because conscious experiences are those which are available to conscious thinkings. Now, although I am in fact one of those who maintain that language mastery is at least contingently connected with the capacity for conscious thought, I shall not argue for this here.[8] Nor is such a thesis necessary in what follows.

[8] See ch. 10 of my *Tractarian Semantics* (New York: Blackwell, 1989).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Are there any other alternatives to my account? I can think of only three. First, it might be said that the distinctive feature of a conscious experience is that it is recorded in short-term memory (this being the *explanation* of why such experiences are, in humans, available to be thought about). But the trouble with this is that there is nothing here to distinguish conscious from nonconscious short-term memory. (My own account, in contrast, is reflexive: conscious thinkings are ones that are themselves available to be consciously thought about.) Second, it might be said that a conscious state is one that is available to the organism as a whole. But the trouble here is that the experiences of any earth worm or slug will turn out to be conscious ones, on this account, whereas the experiences of my car driver are not. Third, it might be claimed that the distinction between conscious and nonconscious states is simple and unanalyzable. But this surely cannot be right. It cannot be merely that we are capable of recognizing, straight off, whether or not a given state is conscious (in the way that we are capable of recognizing whether or not a given shade of color is green); for nonconscious states, precisely because they *are* nonconscious, cannot be immediately recognized as such. Yet if it is said to be the availability for such immediate recognition which constitutes a state as a conscious one, then we appear to have returned to a version of my own proposal.

If my account of the distinction between conscious and nonconscious mental states may be taken as established, then the nonconscious status of most animal experiences follows with very little further argument. For if it is implausible to ascribe second-order beliefs to mice or fish, it is even more unlikely that they should be thinking things consciously to themselves—that is, that they should engage in acts of thinking which are themselves made available for the organism to think about. Indeed, it seems highly implausible to ascribe such activities to any but the higher primates; and, even then, many of us would entertain serious doubts.[9] In the discussion that follows, I shall confine attention to those species for which I take it the above thesis will be noncontroversial. I shall assume that no one would seriously maintain that dogs, cats, sheep, cattle, pigs, or chickens consciously think things to themselves (let alone that fish or reptiles do). In which case, if my account of the distinction between conscious and nonconscious experience is correct, the experiences of all these creatures will be of the nonconscious variety.

[9] See the very sensible discussion of the cognitive powers of chimpanzees in Walker, *op. cit.*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

III.

It goes without saying that pains, too, are experiences. Then two questions remain. First, does pain, like any other experience, admit of a conscious/nonconscious distinction? If so, then the pains of brutes will be nonconscious ones, according to my general account of this distinction. Second, are nonconscious pains an appropriate object of sympathy and moral concern? If not, then the sufferings of brutes will make no moral claims upon us.

There are no noncontroversial examples of nonconscious pain in humans to parallel our everyday examples of nonconscious visual experience. There is an obvious reason for this, since part of the function of pain is to intrude upon consciousness, in order that one may give one's full attention to taking evasive action. But possible examples come from cases where someone is concentrating very intently upon a task, and where they later report having felt no pain upon injury, but where they nevertheless display aversive behavior. For instance, a soldier in the midst of battle may be too caught up in the fighting to notice any pain when he badly burns his hand on the red-hot barrel of a gun, but an observer would see him jerk his hand away in the manner characteristic of someone in pain. Should we feel sympathy in such a case? Clearly we would be sympathetic for the soldier's injury; but not for his suffering, since he in fact felt no pain. This sort of example is incapable of carrying very great weight, however, because the pain behavior displayed is hardly paradigmatic. Since the episode is so brief and unstructured, it may perhaps be thought of as a mere reflex, rather than a genuine instance of nonconscious pain perception.

Can there be cases of pain parallel to those of blindsight? That is, cases where the full (or nearly full) range of pain behavior is displayed, but in which the subject is not conscious of any pain. So far as I am aware, no such cases have actually occurred; but the neurophysiology of pain perception suggests that they are, in principle, possible.[10] Pain in humans is mediated through two types of nerve, which generate distinct projections in the brain subserving distinct functions. Very roughly, the "new path" is fast; it is projected into the higher centers of the brain, and is responsible for precise pain location and fine discriminations of feel. The "old path" is, by contrast, slow; it is projected primarily to the more ancient limbic system in the brain, and gives rise to aversion (the desire for the pain to cease).

[10] In what follows, I rely largely upon the account provided by Dennett in "Why You Can't Make a Computer that Feels Pain," in *Brainstorms*, ch. 10.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Some types of morphine can suppress the activity of the old path, while leaving the new path fully functioning. Patients report that their pain is still just as intense (it feels the same), but that it no longer bothers them (they no longer want it to stop). It seems unlikely, in contrast, that there will be any drug, or any naturally-occurring lesions, which suppress the activity of the new path while leaving the old path functioning. For, unlike the case of vision, the nerves of the new path have no specialized projection area in the higher cortex, but seem rather to be mapped in a complex way into many different regions throughout it.[11] This suggests that phenomena similar to blindsight could only occur as a result of direct surgical intervention. But they do seem to be possible, in principle.

Let us then imagine a case for pain similar to that of blindsight. Suppose that a particular subject, Mary, is never conscious of any pains in her legs. But when she suffers injury in that region, she displays much of normal pain behavior. If we jab pins into her feet, she tends to try very hard to make us stop, she grimaces and groans, and severe damage causes her to scream. But she sincerely declares that she feels nothing. Perhaps she initially found her own behavior disquieting, but now she understands its basis and merely finds it a nuisance. When she twists her ankle, she does not ask for something to alleviate the pain (she says she feels none), but for something to help her relax, and to stop her from grinding her teeth and limping when she walks.

This case is clearly imaginable. It is a possible example (physically possible as well as logically so) of nonconscious pain—that is, of events which otherwise occupy the normal causal role of pain,[12] but which are not available to be thought about consciously and spontaneously by the subject. Ought we to feel sympathy for Mary? We might perhaps feel sympathy for her general condition, since it is in many ways a disturbing situation in which to find oneself. But we should not feel sympathy on specific occasions of injury, since it is clear that she does not suffer. Not being conscious of any pain, her mental state is not an appropriate object of moral concern. (Her

[11] Here I rely upon J. Z. Young, *Philosophy and the Brain* (New York: Oxford, 1986).

[12] This is not strictly correct, since part of the normal causal role of pain is to give rise to a conscious desire that the pain should cease, whereas I am supposing that Mary's avoidance behavior is motivated by desires that are nonconscious. (This is merely for ease of presentation: Mary cannot consciously desire that her *pain* should stop, since she feels none. But she might in fact have a conscious desire that whatever is going on in her ankle should cease.) The adjustment is allowable in this context, since the desires of brutes, as well as their pains, will be nonconscious ones.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

injury itself might be, however, because of its indirect effects upon her life, and hence her conscious desires and disappointments. There are many things that you cannot do with a twisted ankle, even if you feel no pain.) Similarly then in the case of brutes: since their experiences, including their pains, are nonconscious ones, their pains are of no immediate moral concern. Indeed, since all the mental states of brutes are nonconscious, their injuries are lacking even in indirect moral concern. Since the disappointments caused to a dog through possession of a broken leg are themselves nonconscious in their turn, they, too, are not appropriate objects of our sympathy. Hence, neither the pain of the broken leg itself, nor its further effects upon the life of the dog, have any rational claim upon our sympathy.

Of course, it is one thing to reach an intellectual acceptance of such a position, and quite another to put it into practice. Are we really capable of suppressing our sympathy when we see an animal (especially a cuddly one) in severe pain? Not only is it possible that this should occur—after all, the history of mankind is replete with examples of those who have eradicated all feelings of sympathy even for members of other races, by telling themselves that they are not "really human"—but it is a moral imperative that it ought to. Much time and money is presently spent on alleviating the pains of brutes which ought properly to be directed toward human beings, and many are now campaigning to reduce the efficiency of modern farming methods because of the pain caused to the animals involved. If the arguments presented here have been sound, such activities are not only morally unsupportable but morally objectionable.[13]

Consider once again the case of Mary. Suppose that you are a doctor who knows the details of her condition, and that you happen to be on the scene of an accident in which her legs have been badly injured. A number of other people are obviously injured and in pain, but Mary is screaming the loudest. Ought you to treat her first? Clearly not, other things being equal (e.g., provided that she is not bleeding badly); indeed, it would be moral weakness in you to do so. For you know that she is not really suffering, since her pains are

[13] Peter Singer has been prominent in defending the moral significance of animals, from a broadly utilitarian perspective. See, for example, his *Practical Ethics* (New York: Cambridge, 1979). But he makes no attempt to take account of the distinction between conscious and nonconscious experience. Indeed, he does not notice it. There are, of course, other moral perspectives, from the standpoint of some of which the moral significance of animals may be attacked; for example, versions of contractualism which place a premium upon rational agency. But my argument is the more fundamental, since even contractualists should find a place for compassion.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

nonconscious, whereas the sufferings of the others are real. Similarly then in the case of brutes: since their pains are nonconscious (as are all their mental states), they ought not to be allowed to get in the way of any morally–serious objective.

It is worth drawing a contrast at this point with the case of very young children. I presume that the pains of babies, too, are nonconscious; for no one will seriously maintain that they consciously think things to themselves. Nevertheless, it is important that they should continue to evoke our sympathy. For a baby's pains and injuries, and our attitudes toward them, are likely to have a significant effect upon the person the baby will one day become. There is a parallel here with the case of language development. As every parent knows, one naturally becomes possessed by a sort of necessary insanity, talking to young children as if they could understand. This is no doubt essential if children are to develop into skilled practitioners of their native tongue. In both cases, one engages, and should engage, in a useful fiction: ascribing to the baby thoughts that it does not in fact possess.

There is no such rationale in the case of brutes. For our sympathy and concern for their pains and injuries cannot be said to have an effect on the persons they will one day become. Nevertheless, it is hard, especially in a culture such as ours where animals are often used as exemplars in moral training (we tell children that it is *cruel* to pull the whiskers out of the cat), to eradicate our feelings toward them. Indeed, in such a culture we may have reason to look askance at people who can be wholly indifferent to an animal writhing in agony before their very eyes. (We say, "How can you be so inhuman?") But what should be said is that we are obliged to set such feelings aside whenever they threaten to have a morally significant effect upon other humans. And it also follows that there is no moral criticism to be leveled at the majority of people who are indifferent to the pains of factory-farmed animals, which they know to exist but do not themselves observe.

PETER CARRUTHERS

University of Essex

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

BOOK REVIEWS

Philosophical Papers, 2 vols. CHARLES TAYLOR. New York: Cambridge University Press, 1985. 294 p. and 340 p. Each volume: cloth $39.50, paper $12.95.

"When a man tells me that he is neither of the Right nor of the Left," wrote the French political commentator, Alain, "then I know that I am talking to a man of the Right." It is easy to feel this way about analytical philosophy. Was there ever a philosophical tradition with a less articulated account of its own historical distinctiveness?

Recently, however, there have been a number of attempts to throw the background assumptions of analytical philosophy into relief by examining it from a historical perspective. One thinks, in particular, of the work of Alasdair MacIntyre and Richard Rorty, as well as Charles Taylor, whose *Philosophical Papers* are here under review. What these undertakings have in common is that they aim to pass from the level of particular arguments to a historical standpoint from which the arguments can themselves be displayed as embodying assumptions and preconceptions taken over unreflectively from the wider culture. All three authors would agree with Hegel's view of philosophy as "its own time captured in thought." In each case, a certain central area of philosophy, it is claimed, is so far penetrated by the prejudices of a scientific culture as to undermine the value of the entire enterprise.

What, though, is the status of such criticism? Is it that philosophical arguments are assessed and shown to be inadequate in their own terms and according to their own standards? If so, then the point of moving beyond the terrain of philosophical dispute as conventionally understood is not to inquire whether the beliefs in question are true, but to ask a further, distinct question: why, given that such beliefs are *not* true, do people nevertheless continue to hold them? From this point of view, the historical approach does not represent a challenge to the received conception of philosophical method, but amounts to the conjunction of two separate enterprises: philosophy and (for want of a better word) the sociology of knowledge.

On a more radical view, however, the two aspects cannot be separated: a strict division between questions concerning the origins and motives of beliefs, on the one hand, and questions regarding their validity, on the other, is itself part of the Enlightenment tradition which is being called into question. Historical considerations may themselves form part of the argument as to whether to hold a partic-

0022-352X/89/8605/270–276 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ular philosophical view or not. In this case, the methodological status of the criticism does, indeed, become problematic. For what would count as a reason to accept *that* as a procedure for philosophical argument? Would it, too, be historical in character?

In what follows, I shall show that the position advocated by Taylor in his *Philosophical Papers* does, indeed, depend on such a broad conception of philosophical inquiry. This is not to say that Taylor's position is invalid (or, indeed, unattractive), but that the arguments it presents fall short of what would be ideally conclusive—the strict entailment of his position by reasons that all parties would acknowledge as compelling. Taylor would argue that such an ideal of proof is part of the Enlightenment view of philosophy. In fact, for philosophy, "[as] for any hermeneutic explanation, interpretive plausibility is the ultimate criterion" (vol. I, p. 7).

I am broadly in agreement with Taylor on this issue, but it opens up a possibility that I believe he does not sufficiently consider: that more than one interpretation may turn out to be finally plausible. If so, then Taylor and his opponents may face a kind of "Mexican stand-off"; each position—though clearly mutually incompatible—may be sufficiently supported by internal reasons as to appear rational to those who hold it, while being unable to disarm the other on the other's own ground.

The subject matter covered in the *Philosophical Papers* ranges from the philosophy of mind and action through to the philosophy of social science and political theory. It is impossible to overpraise the manner in which it is presented. Each essay is clear, vigorous, vivid, and unpretentious. They are the writings of a man whose conviction of the merits of what he has to say is so overwhelming that he is happy—indeed, eager—to place his cards as plainly as possible on the table.

What unifies this apparently disparate material is the attack on a doctrine that Taylor calls "naturalism." Naturalism is not simply the general view that man can be seen as part of nature—that, Taylor agrees, would be accepted in some sense or other by everyone—but the more specific belief, he claims, that the nature of which man is part is "to be understood according to the canons which emerged in the seventeenth-century revolution in natural science" (vol. I, p. 2).

Naturalism has dominated analytical philosophy, Taylor believes. Yet naturalism, he claims, "cannot cope with our understanding of the self"; there are aspects of that understanding—chiefly, that we are self-interpreting beings for whom things have significance in a strong sense—which "cannot be captured by a scientific language which essentially aspires to neutrality" (vol. I, pp. 3/4).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

As represented by Taylor, then, the central core of naturalism is its adhesion to the canons of seventeenth–century natural science. But what this amounts to is difficult to answer, for Taylor's interpretive strategy proceeds on two levels. At one level, a crude and obviously unsatisfactory view is presented which (as Taylor himself would no doubt admit) no naturalist would accept as an accurate representation of his position. The claim, however, is that the crude form of the view is something like an ideal type—it expresses the unacknowledged source of the more sophisticated versions of the view's appeal and forms a model from which actual theories depart only at the price of a loss of content. Utilitarianism is an example. Not all utilitarians are simple hedonists. But, to the extent that utilitarianism departs from that naive basic model, so, Taylor believes, utilitarianism loses the connection to the deep source of its own appeal: the idea of an objectively calculable resolution of all moral dilemmas.

In their crudest form, one supposes, the canons of seventeenth-century natural science might amount to something like this: what is real in the highest degree is matter in motion and true explanations must reduce to the mathematical laws governing that. No one now takes such a simplistic position (perhaps, indeed, no one ever did), but it inspires, Taylor claims, two related beliefs. First, there is an ontological doctrine—a desire to account for the world as composed solely of what Taylor calls "non-anthropocentric" properties. Second, there is a reductive impulse in favor of causal explanations framed in terms of such properties.

How might the naturalist respond to this? It is very hard to defend oneself against attack at such a high level of generality. Taylor does not accuse the naturalist of believing in any particular list of ultimate properties, but merely that whatever properties are to be ultimate must be "non-anthropocentric"—a characterization which is itself left rather unclear. Nor will there be many naturalists who believe that all higher-order theories can be eliminated in favor of some single basic level of explanation. But, if not this, then to what does the charge of "reductionism" amount? For such reasons, there is some temptation to dismiss Taylor's whole enterprise as conducted at a level too vague to engage with the specific contours of the theories under attack.

Yet it is also clear why Taylor is forced to present his argument in the way he does. If his case is to have the kind of generality to which he aspires, it must be framed, not at the level of the specific doctrines that take the foreground in debate, but at the level of the back-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ground beliefs that motivate them. A more powerful reply would take up Taylor's challenge at this level.

Naturalism is, Taylor says, "manifestly reductive"; just as utilitarianism is inspired by hedonism, so behaviorism is the ideal form of naturalism:

> . . . the ideally simple theory of this objective kind would be a behaviourist one, where we could correlate physical features of the environment with movements of the organism; and this helps account for the continuing obsession with this wildly implausible approach to the science of man (vol. I, p. 51).

But surely, the naturalist will object, it is a gross caricature to depict naturalists as closet behaviorists, the more so when the explosion of modern cognitive science—psychological functionalism, AI research, computational linguistics, and so on—seems to point in just the opposite direction: that physicalist accounts of complex goal-directed systems do not (and cannot) aim toward the elimination of higher-order predicates in favor of simple physical ones.

Taylor takes up this crucial issue in a discussion of Jerry Fodor's *The Language of Thought*. Since Fodor explicitly defends the irreducibility of psychology to physics, he would, it would appear, provide a counterexample to Taylor's view that modern naturalism is reductionist in motivation. But this is not so, Taylor believes. Although cognitive psychology does not aspire to an explanation that eliminates everything but the physical features of the environment and the motions of the organism within it, reductionism comes in at an earlier stage: in the move from our own account of ourselves as agents to a scientifically-oriented psychology that accounts for what we do in computational terms. The issue, then, is not the elimination of features that reflect significance and interpretation as such, but whether the equivalent of such features found in cognitive psychology forms an adequate starting point for understanding them in the human context. Taylor believes that the supposition is absurd:

> What we normally understand as the predicates of psychology . . . plainly do not apply to machines. Nor do we have anything but the vaguest fantasies as to how they might apply in the future (vol. I, pp. 209–210).

Human beings, Taylor claims, possess a "significance feature" by which:

> . . . in contradistinction to machines, we attribute action to ourselves in a strong sense, a sense in which there is an answer to the question, What

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

is he doing? which is not merely relative to the interests and purposes of an observer (vol. I, p. 194).

Taylor's case, then, is that the significance that cognitive psychology apparently concedes is not *real* significance—and what differentiates true significance from pseudo significance is that we ourselves are the ultimate source of authority as to our own purposes. At this point, the naturalist may perhaps feel that he has treed his quarry. Either Taylor's position is purely stipulative (only human beings can have true, human purposes) or it appears to relapse into a form of Cartesianism. What else but the idea that the inner mental state of the agent has final authority could make one suppose that there is a matter of fact distinguishing between what are merely "adequate" explanations of an action and an account of what the agent is *really* doing?

But, of course, Taylor will not find such a counter compelling, any more than the naturalist will be convinced by what he will regard as arm-waving sermonizing about true human significance. Rather than pursue a futile combat, the naturalist may want to take a step back and place it in historical context. For have we not been here before? Since at least the eighteenth century, philosophers concerned for the dignity of human nature have been trying to hold back the tide of scientific enlightenment. At that time, it was organic nature whose purposiveness and creativity would, it was supposed, close it forever to the rigidities of mere causal explanation—it was no less than Kant who told us that there would never be a "Newton" to explain a blade of grass. Surely there is something absurd at the spectacle of philosophers taking up supposedly final positions against science, only to have to give them up as soon as the water starts to lap around their ankles.

Taylor is not impressed by such considerations. What they amount to, he says, is "less an argument from logical necessity than a persuasive use of historical analogy" (vol. I, p. 180). Indeed so—but Taylor can hardly object to that. Argument from history may not be conclusive, but that does not mean that it is valueless.

Yet the argument from past success can be balanced by an argument from present failure. For when the methods of naturalism have been applied to the study of man in the form of behaviorism and positivist "political science," the results have been laughable—empirical banalities and empty generalizations dressed up in pompous pseudo-scientific form. Science has, it is true, often over extended itself in the past (we should not let the example of phrenology dis-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

credit the case for neuro-physiology), but Taylor's powerful and often highly entertaining demolition of behaviorist and positivist trends cannot be discounted. Taylor might be wrong in seeing behaviorism as the essential expression of naturalism in the human sciences, but he is certainly right about its inadequacies.

If the argument from historical progress cannot settle matters, then the only further possibility is to turn to the underlying motivation behind naturalism. Although he often talks as if he had refuted naturalism in its own terms, Taylor appreciates that naturalists do not accept his arguments as conclusive:

> . . . the long experience of polemic against naturalist views on the epistemological level of the philosophy of science, the sense of futility when one fails to carry conviction against what seem ultimately absurd views, convinces me that the real issue lies elsewhere (vol. I, p. 6).

It lies, in the end, he claims, in naturalism's moral appeal, insofar as it is sustained by a certain attractive picture of human agency, that of the disengaged identity, free from the limiting constraints of empirical contingencies and traditions. To show this would be the final refutation of naturalism:

> . . . to show that its opponents understand it better than it does itself, that indeed the phenomenon of people believing in naturalism was only explicable in terms of a rival incompatible theory (vol. I, p. 6).

But it is just here that Taylor's argument is weakest. For although he can certainly claim to have shown that one is not compelled to accept naturalism as a method in the human sciences on grounds of reasonableness alone, his attempts to construct a historical interpretation of the development of the modern identity are simplistic and misleading.

Like so many philosophers who take inspiration from the Continental tradition, Taylor treats Cartesianism as the ideal type of modern philosophy. An ontology of the world as extended matter is coupled (quite mysteriously) to an isolated but self-certain conscious agent. It is the appeal of the latter which, it is alleged, sustains belief in the former.

But to treat Descartes as the founder of the naturalist view of agency is quite wrong. What it neglects is the degree to which Cartesianism was itself part of a defensive maneuver. An important element in Cartesianism was the attempt to accommodate modern phys-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ics, as far as possible, to a traditional theological view of agency: the idea of human beings as sufficiently free in the origination of their actions as to be able to bear the full weight of moral responsibility for them. Insofar as the disengaged identity has been a powerful ingredient in modern consciousness, it is this which motivates it.

To regard the disengaged identity as the fundamental motivation behind modern naturalism is to ignore, however, the extent to which naturalism has aimed to dissolve precisely that Christian notion of agency and identity. In that respect, the founder of naturalism is not Descartes but Hume (and, perhaps, Hobbes) and his rejection of moral rationalism and radical freedom. To put this together with Descartes's dualistic ontology as ingredients in a single conception of the self, as Taylor does, is to coalesce what are, in fact, fundamentally opposed endeavors.

Nor is it by any means true, the naturalist will argue, that Taylor can explain the naturalist's beliefs in a way that the naturalist himself cannot. The Weberian, to take an example, is only too well able to understand why the process of secularization should produce, not only the naturalist belief in the indefinite extensibility of the principles of science, but also the idealist longing for the consolations of belief in human "specialness."

Taylor's interpretation will, in short, be opposed by the naturalist at every turn. I have tried in this review to delineate that position as clearly as possible in limited space. But this is not intended as a refutation of Taylor, much less as a vindication of naturalism itself. I simply want to argue that the naturalist's position is not unreasonable. Taylor himself remains entirely deaf to naturalism's appeal. In the end, he stands outside it as a bemused (and slightly exasperated) anthropologist. What I have tried to suggest is that this is not necessarily the final standpoint: that a perspective exists from which it is possible, consistently, to feel the force of both claims.

MICHAEL E. ROSEN

University College/London

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

NEW BOOKS

ABELSON, RAZIEL: *Lawless Mind.* Philadelphia: Temple, 1988. xix, 209 p. Cloth $29.95.

AGASSI, JOSEPH: *The Gentle Art of Philosophical Polemics.* Peru, IL: Open Court, 1988. 521 p. Cloth $49.95, paper $19.95.

BAARS, BERNARD J.: *A Cognitive Theory of Consciousness.* New York: Cambridge, 1988. xxiii, 424 p. Cloth $44.50.

BAKER, GORDON: *Wittgenstein, Frege & the Vienna Circle.* New York: Blackwell, 1988. xxii, 274 p. Cloth $34.95.

BORUAH, BIJOY H.: *Fiction and Emotion: A Study in Aesthetics and the Philosophy of Mind.* New York: Oxford, 1988. viii, 133 p. Cloth $36.00.

BOSTOCK, DAVID: *Plato's* Theaetetus. New York: Oxford, 1988. viii, 285 p. Cloth $55.00.

BRENNAN, ANDREW: *Thinking about Nature: An Investigation of Nature, Value, and Ecology.* Athens: Georgia, 1988. xiii, 235 p. Cloth $30.00, paper $15.00.

BRINK, DAVID O.: *Moral Realism and the Foundation of Ethics.* New York: Cambridge, 1989. xii, 340 p. Cloth $44.50, paper $16.95.

CADY, DUANE L.: *From Warism to Pacifism: A Moral Continuum.* Philadelphia: Temple, 1989. xvii, 159 p. Cloth $19.95.

CALLICOT, J. BAIRD: *In Defense of the Land Ethic: Essays in Environmental Philosophy.* Albany: SUNY, 1989. x, 325 p. Cloth $35.00, paper $10.95.

CAMPBELL, ROBERT, and DIANE COLLINSON: *Ending Lives.* New York: Blackwell, 1988. xviii, 203 p. Cloth $29.95, paper $11.95.

CASSIRER, HEINZ W.: *Grace & Law: St. Paul, Kant, and the Hebrew Prophets.* Grand Rapids, MI: Eerdmans, 1988. xvi, 176 p. Cloth $12.95.

CLARKE, D. S., JR.: *Rational Acceptance and Purpose: An Outline of a Pragmatist Epistemology.* Totowa, NJ: Rowman & Littlefield, 1989. x, 148 p. Cloth $29.50, paper $14.95.

COLLINS, STEPHEN L.: *From Divine Cosmos to Sovereign State.* New York: Oxford, 1989. viii, 235 p. Cloth $39.95.

COPPLESTON, FREDERICK C.: *Russian Religious Philosophy: Selected Aspects.* Notre Dame: University Press, 1988. xii, 158 p. Cloth $24.95.

CULLER, JONATHAN: *Framing the Sign: Criticism and its Institutions.* Norman: Oklahoma, 1989. xviii, 237 p. Cloth $29.95.

D'AMICO, ROBERT: *Historicism and Knowledge.* New York: Routledge, 1989. xiv, 174 p. Cloth $35.00, paper $12.95.

DETMER, DAVID: *Freedom as a Value: A Critique of the Ethical Theory of Jean-Paul Sartre.* Peru, IL: Open Court, 1988. ix, 262 p. Cloth $29.95, paper $15.95.

DONAGAN, ALAN: *Spinoza.* Chicago: University Press, 1989. xviii, 219 p. Cloth $34.95.

FLATHMAN, RICHARD E.: *Toward a Liberalism.* Ithaca: Cornell, 1989. 223 p. Cloth $29.95, paper $10.95.

GARDINER, PATRICK: *Kierkegaard.* New York: Oxford, 1988. 120 p. Cloth $29.95, paper $5.95.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

GILLIS, CHESTER: *A Question of Final Belief: John Hick's Pluralistic Theory of Salvation.* New York: St. Martin's Press, 1989. xiii, 186 p. Cloth $39.95.

GOLDMAN, ALAN H.: *Moral Knowledge.* New York: Routledge, 1988. ix, 224 p. Cloth $35.00.

GORDON, WENDELL, and JOHN ADAMS: *Economics as a Social Science: An Evolutionary Approach.* Riverdale, MD: Riverdale, 1989. ix, 254 p. Cloth $33.00, paper $25.00.

GRAHAM, GORDON: *Contemporary Social Philosophy.* New York: Blackwell, 1988. viii, 185 p. Cloth $45.00, paper $16.95.

GREEN, GARRETT: *Imagining God: Theology & the Religious Imagination.* San Francisco: Harper & Row, 1989. x, 179 p. Cloth $22.95.

HALEY, MICHAEL CABOT: *The Semeiosis of Poetic Metaphor.* Bloomington: Indiana UP, 1988. xiv, 178 p. Cloth $25.00.

HAZELRIGG, LAWRENCE: *Claims of Knowledge: On the Labor of Making Found Worlds.* Tallahassee: Florida State UP, 1989. v, 317 p. Cloth $25.00.

HAZELRIGG, LAWRENCE: *A Wilderness of Mirrors: On Practices of Theory in a Gray Age.* Tallahassee: Florida State UP, 1989. viii, 520 p. Cloth $38.00.

HELLER, ERICH: *The Importance of Nietzsche: Ten Essays.* Chicago: University Press, 1989. xi, 199 p. Cloth $29.95, paper $11.95.

HOLMES, ROBERT L.: *On War and Morality.* Princeton: University Press, 1989. xiii, 310 p. Cloth $35.00, paper $13.95.

HORN, LAURENCE R.: *A Natural History of Negation.* Chicago: University Press, 1989. xxii, 637 p. Cloth $75.00, paper $34.95.

IRWIN, TERENCE: *Classical Thought.* New York: Oxford, 1989. xii, 266 p. Cloth $29.95, paper $10.95.

JOHNSON, MARK: *Attributive-Value Logic and the Theory of Grammar.* Chicago: University Press, 1988. xi, 167 p. Cloth $37.50, paper $15.95.

KEARNEY, RICHARD: *The Wake of the Imagination.* Minneapolis: Minnesota UP, 1988. 467 p. Cloth $45.00, paper $19.95.

KIRWIN, CHRISTOPHER: *Augustine.* New York: Routledge, 1989. viii, 247 p. Cloth $75.00.

LANGHOLM, TORE: *Partiality, Truth and Persistence.* Chicago: University Press, 1989. viii, 151 p. Cloth $29.95, paper $12.95.

LEVINE, MICHAEL P.: *Hume and the Problem of Miracles: A Solution.* Boston: Kluwer, 1989. viii, 212 p. Cloth $64.00.

LITTLE, J. P.: *Simone Weil: Waiting on Truth.* New York: St. Martin's, 1988. 170 p. Cloth $21.95.

LLOYD, ELIZABETH A.: *The Structure and Confirmation of Evolutionary Theory.* Westport, CT: Greenwood, 1988. viii, 235 p. Cloth $37.95.

LUBAN, DAVID: *Lawyers and Justice: An Ethical Study.* Princeton: University Press, 1989. xxix, 440 p. Cloth $55.00, paper $12.95.

MADISON, G. B.: *The Hermeneutics of Postmodernity: Figures and Themes.* Bloomington: Indiana UP, 1988. xvi, 206 p. Cloth $35.00.

MARENBON, JOHN: *Early Medieval Philosophy (480–1150): An Introduction.* New York: Routledge, 1988. xv, 197 p. Paper $13.95.

MARSH, JAMES L.: *Post-Cartesian Meditations.* New York: Fordham, 1989. xiii, 279 p. Cloth $40.00.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

McCUMBER, JOHN: *Poetic Interaction.* Chicago: University Press, 1989. xiv, 489 p. Cloth $60.00, paper $24.95.

McKNIGHT, STEPHEN A.: *Sacralizing the Secular: The Renaissance Origins of Modernity.* Baton Rouge: Louisiana UP, 1989. xi, 131 p. Cloth $25.00.

MELLEMA, GREGORY: *Individuals, Groups, and Shared Moral Responsibility.* New York: Lang, 1988. 215 p. Cloth $33.50.

MITSIS, PHILLIP: *Epicurus' Ethical Theory: The Pleasures of Invulnerability.* Ithaca: Cornell, 1988. ix, 184 p. Cloth $24.95.

MOLNAR, THOMAS: *Twin Powers: Politics and the Sacred.* Grand Rapids, MI: Eerdmans, 1988. xi, 147 p. Paper $9.95.

MOORE, HENRIETTA: *Feminism and Anthropology.* Minneapolis: Minnesota UP, 1988. ix, 246 p. Cloth $39.50, paper $15.95.

MURPHY, JEFFRIE G., and JEAN HAMPTON: *Forgiveness and Mercy.* New York: Cambridge, 1988. xii, 194 p. Cloth $29.95.

OAKES, GUY: *Weber and Rickert: Concept Formation in the Cultural Sciences.* Cambridge: MIT, 1989. viii, 190 p. Cloth $22.50.

PAEHLKE, ROBERT C.: *Environmentalism and the Future of Progressive Politics.* New Haven: Yale, 1989. viii, 325 p. Cloth $25.00.

PANAYOT, BUTCHVAROV: *Skepticism in Ethics:* Bloomington: Indiana UP, 1989. viii, 225 p. Cloth $37.50, paper $14.95.

PEARS, DAVID: *The False Prison: A Study of the Development of Wittgenstein's Philosophy.* New York: Oxford, 1988. ix, 341 p. Cloth $39.95, paper $16.95.

PERCESEPE, GARY JOHN: *Future(s) of Philosophy: The Marginal Thinking of Jacques Derrida.* New York: Lang, 1989. 227 p. Cloth $36.00.

PEREZ-RAMOS, ANTONIO: *Francis Bacon's Idea of Science and the Maker's Knowledge Tradition.* New York: Oxford, 1988. xi, 334 p. Cloth $77.00.

RADDER, HANS: *The Material Realization of Science: A Philosophical View on the Experimental Natural Sciences, Developed in Discussion with Habermas.* Assen: Van Gorcum, 1988. viii, 204 p.

RADEST, HOWARD B.: *Can We Teach Ethics?* New York: Praeger, 1989. xii, 168 p. Cloth $39.95.

RAMSEY, ALLAN: *Formal Methods in Artificial Intelligence.* New York: Cambridge, 1988. ix, 279 p. Cloth $49.50.

READ, STEPHEN: *Relevant Logic.* New York: Blackwell, 1988. vii, 199 p. Cloth $45.00.

RESCHER, NICHOLAS: *Rationality: A Philosophical Inquiry into the Nature and the Rationale of Reason.* New York: Oxford, 1988. xi, 234 p. Cloth $45.00.

RILEY, DENISE: *Am I That Name?: Feminism and The Category of Women in History.* Minneapolis: Minnesota UP, 1989. vi, 126 p. Cloth $29.95, paper $12.95.

RORTY, RICHARD: *Contingency, Irony, and Solidarity.* New York: Cambridge, 1989. xvi, 201 p. Cloth $34.50.

ROSEN, STANLEY: *The Ancients and the Moderns: Rethinking Modernity.* New Haven: Yale, 1989. x, 236 p. Cloth $22.50.

RUDDICK, SARA: *Maternal Thinking: Toward a Politics of Peace.* Boston: Beacon, 1989. xi, 291 p. Cloth $24.95.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

RUNCIMAN, W. G.: *A Treatise on Social Theory: vol. II: Substantive Social Theory.* New York: Cambridge, 1989. xiii, 493 p. Cloth $59.50, paper $18.95.

SAMUELSON, NORBERT M.: *An Introduction to Modern Jewish Philosophy.* Albany: SUNY, 1989. x, 320 p. Cloth $49.50, paper $14.95.

SANDERS, ANDY F.: *Michael Polanyi's Post-Critical Epistemology: A Reconstruction of Some Aspects of "Tacit Knowing."* Amsterdam: Rodopoi, 1988. iv, 294 p.

SCHOULS, PETER A.: *Descartes and the Enlightenment.* Montreal: McGill-Queen's, 1989. 194 p.

SILVERS, STUART: *Rerepresentation: Readings in the Philosophy of Mental Representation.* Boston: Kluwer, 1989. xxiii, 417 p. Cloth $79.00.

SIRCELLO, GUY: *Love and Beauty.* Princeton: University Press, 1989. vi, 253 p. Cloth $26.50.

SPELMAN, ELIZABETH V.: *Inessential Woman: Problems of Exclusion in Feminist Thought.* Boston: Beacon, 1989. xiii, 221 p. Cloth $22.95.

TRIGG, ROBERT: *Ideas of Human Nature: An Historical Introduction.* New York: Blackwell, 1988. 184 p. Cloth $39.95, paper $16.95.

WEISS, PAUL: *Philosophy in Process, vol. Eleven: 19 January 1986–27 May 1987.* Albany: SUNY, 1989. 399 p. Cloth $76.95, paper $24.50.

WINN, JANET: *The Open Mind and Sociology.* New York: Lang, 1988. 197 p. Cloth $34.50.

ZIMMERMAN, MICHAEL J.: *An Essay on Moral Responsibility.* Totowa, NJ: Rowman & Littlefield, 1988. xiii, 210 p. Cloth $38.50.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVI, NO. 6, JUNE 1989

BIOSEMANTICS

Causal or informational theories of the semantic content of mental states which have had an eye on the problem of false representations have characteristically begun with something like this intuition. There are some circumstances under which an inner representation has its represented as a necessary and/or sufficient cause or condition of production. That is how the content of the representation is fixed. False representations are to be explained as tokens that are produced under other circumstances. The challenge, then, is to tell what defines certain circumstances as the content–fixing ones.

I.

Note that the answer cannot be just that these circumstances are *statistically* normal conditions. To gather such statistics, one would need to delimit a reference class of occasions, know how to count its members, and specify description categories. It would not do, for example, just to average over conditions-in-the-universe-any-place-any-time. Nor is it given how to carve out relevant description categories for conditions on occasions. Is it "average" in the summer for it to be (precisely) between 80° and 80.5° Fahrenheit with humidity 87%? And are average conditions those which obtain on at least 50% of the occasions, or is it 90%? Depending on how one sets these parameters, radically different conditions are "statistically normal." But the notion of semantic content clearly is not relative, in this manner, to arbitrary parameters. The content-fixing circumstances must be *nonarbitrarily* determined.

A number of recent writers have made an appeal to teleology here, specifically to conditions of normal function or well-functioning of the systems that produce inner representations. Where the represented is R and its representation is "R," under conditions of well-functioning, we might suppose, only Rs can or are likely to produce

0022-362X/89/8606/281–297 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

"Rs." Or perhaps "R" is a representation of *R* just in case the system was designed to react to *R*s by producing "Rs." But this sort of move yields too many representations. Every state of every functional system has normal causes, things that it is a response to in accordance with design. These causes may be proximate or remote, and many are disjunctive. Thus, a proximate normal cause of dilation of the skin capillaries is certain substances in the blood, more remote causes include muscular effort, sunburn, and being in an overheated environment. To each of these causes the vascular system responds by design, yet the response (a red face), though it may be a natural sign of burn or exertion or overheating, certainly is not a representation of that. If not every state of a system represents its normal causes, which are the states that do?

Jerry Fodor[1] has said that, whereas the content of an inner representation is determined by some sort of causal story, its status *as* a representation is determined by the functional organization of the part of the system which uses it. There is such a thing, it seems, as behaving like a representation without behaving like a representation of anything in particular. What the thing is a representation of is then determined by its cause under content-fixing conditions. It would be interesting to have the character of universal I-am-a-representation behavior spelled out for us. Yet, as Fodor well knows, there would still be the problem of demonstrating that there was only one normal cause per representation type.

A number of writers, including Dennis Stampe,[2] Fred Dretske,[3] and Mohan Matthen,[4] have suggested that what is different about effects that are representations is that their function is, precisely, to represent, "indicate," or "detect." For example, Matthen says of (fullfledged) perceptual states that they are "state[s] that [have] the function of *detecting* the presence of things of a certain type . . ." (*ibid.*, p. 20). It does not help to be told that inner representations are things that have representing (indicating, detecting) as their function, however, unless we are also told what kind of activity repre-

[1] "Banish Discontent," in Jeremy Butterfield, ed., *Language, Mind and Logic* (New York: Cambridge, 1986), pp. 1–23; *Psychosemantics: The Problem of Meaning in the Philosophy of Mind* (Cambridge: MIT, 1987).

[2] "Toward a Causal Theory of Representation," in Peter French, Theodore Uehling Jr., Howard Wettstein, eds., *Contemporary Perspectives in the Philosophy of Language* (Minneapolis: Minnesota UP, 1979), pp. 81–102.

[3] "Misrepresentation," in Radu Bogdan, ed., *Belief: Form, Content, and Function* (New York: Oxford, 1986), pp. 17–36.

[4] "Biological Functions and Perceptual Content," this JOURNAL, LXXXV, 1 (January 1988):5–27.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

senting (indicating, detecting) is. Matthen does not tell us how to naturalize the notion "detecting." If "detecting" is a function of a representational state, it must be something that the state effects or produces. For example, it cannot be the function of a state to have been produced in response to something. Or does Matthen mean that it is not the representational states themselves, but the part of the system which produces them, which has the function of detecting? It has the function, say, of producing states that correspond to or covary with something in the outside world? But, unfortunately, not every device whose job description includes producing items that vary with the world is a representation producer. The devices in me that produce calluses are supposed to vary their placement according to where the friction is, but calluses are not representations. The pigment arrangers in the skin of a chameleon, the function of which is to vary the chameleon's color with what it sits on, are not representation producers.

Stampe and Dretske do address the question what representing or (Dretske) "detecting" is. Each brings in his own description of what a natural sign or natural representation is, then assimilates *having the function of representing R* to being a natural sign or representer of *R* when the system functions normally. Now, the production of natural signs is undoubtedly an accidental side effect of normal operation of many systems. From my red face you can tell that either I have been exerting myself, or I have been in the heat, or I am burned. But the production of an accidental side effect, no matter how regular, is not one of a system's functions; that goes by definition. More damaging, however, it simply is not true that representations must carry natural information. Consider the signals with which various animals signal danger. Nature knows that it is better to err on the side of caution, and it is likely that many of these signs occur more often in the absence than in the presence of any real danger. Certainly there is nothing incoherent in the idea that this might be so, hence that many of these signals do not carry natural information concerning the dangers they signal.

II.

I fully agree, however, that an appeal to teleology, to function, is what is needed to fly a naturalist theory of content. Moreover, what makes a thing into an inner representation is, near enough, that its function is to represent. But, I shall argue, the way to unpack this insight is to focus on representation *consumption,* rather than representation production. It is the devices that *use* representations which determine these to be representations and, at the same time (contra

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Fodor), determine their content. If it really is the function of an inner representation to indicate its represented, clearly it is not just a natural sign, a sign that you or I looking on might interpret. It must be one that functions as a sign or representation *for the system itself.* What is it then for a system to use a representation *as* a representation?

The conception of function on which I shall rely was defined in my *Language, Thought, and Other Biological Categories*[5] and defended in "In Defense of Proper Functions"[6] under the label "proper function." Proper functions are determined by the histories of the items possessing them; functions that were "selected for" are paradigm cases.[7] The notions "function" and "design" should not be read, however, as referring only to origin. Natural selection does not slack after the emergence of a structure but actively preserves it by acting against the later emergence of less fit structures. And structures can be preserved due to performance of new functions unrelated to the forces that originally shaped them. Such functions are "proper functions," too, and are "performed in accordance with design."

The notion "design" should not be read—and this is very important—as a reference to innateness. A system may have been designed to be altered by its experience, perhaps to learn from its experience in a prescribed manner. Doing what it has learned to do in this manner is then "behaving in accordance with design" or "functioning properly."[8]

My term 'normal' should be read normatively, historically, and relative to specific function. In the first instance, 'normal' applies to explanations. A "normal explanation" explains the performance of a particular function, telling how it was (typically) historically performed on those (perhaps rare) occasions when it was properly performed. Normal explanations do not tell, say, why it has been common for a function to be performed; they are not statistical explanations. They cover only past times of actual performance, showing how these performances were entailed by natural law, given certain conditions, coupled with the dispositions and structures of the rele-

[5] Cambridge: MIT, 1984 (hereafter LTOBC).

[6] *Philosophy of Science,* LVI, 2 (June 1989): 288–302.

[7] An odd custom exists of identifying this sort of view with Larry Wright, who does not hold it. See my "In Defense of Proper Functions." Natural selection is not the only source of proper functions. See LTOBC, chs. 1 and 2.

[8] See LTOBC; and "Truth Rules, Hoverflies, and the Kripke-Wittgenstein Paradox," *The Philosophical Review* (forthcoming).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

vant functional devices.[9] In the second instance, 'normal' applies to conditions. A "normal condition for performance of a function" is a condition, the presence of which must be mentioned in giving a full normal explanation for performance of that function. Other functions of the same organism or system may have other normal conditions. For example, normal conditions for discriminating colors are not the same as normal conditions for discriminating tastes, and normal conditions for seeing very large objects are not the same as for seeing very small ones. It follows that 'normal conditions' must not be read as having anything to do with what is typical or average or even, in many cases, at all common. First, many functions are performed only rarely. For example, very few wild seeds land in conditions normal for their growth and development, and the protective colorings of caterpillars seldom actually succeed in preventing them from being eaten. Indeed, normal conditions might almost better be called "historically optimal" conditions. (If normal conditions for proper functioning, hence survival and proliferation, were a statistical norm, imagine how many rabbits there would be in the world.) Second, many proper functions only need to be performed under rare conditions. Consider, for example, the vomiting reflex, the function of which is to prevent (further) toxification of the body. A normal condition for performance of this function is presence, specifically of poison in the stomach, for (I am guessing) it is only under that condition that this reflex has historically had beneficial effects. But poison in the stomach certainly is not an average condition. (Nor, of course, is it a normal condition for other functions of the digestive system.[10])

If it is actually one of a system's functions to produce representations, as we have said, these representations must function as representations for the system itself. Let us view the system, then, as divided into two parts or two aspects, one of which produces representations for the other to consume. What we need to look at is the consumer part, at what it is to use a thing *as* a representation.

[9] This last clarification is offered to aid Fodor ("On There Not Being an Evolutionary Theory of Content" [hereafter NETC], forthcoming), who uses my term 'Normal' (here I am not capitalizing it but the idea has not changed) in a multiply confused way, making a parody of my views on representation. In this connection, see also fns. 13 and 17.

[10] "Normal explanation" and "normal condition for performance of a function," along with "proper function," are defined with considerable detail in LTOBC. The reader may wish, in particular, to consult the discussion of normal explanations for performance of "adapted and derived proper functions" in ch. 2, for these functions cover the functions of states of the nervous system which result in part from learning, such as states of human belief and desire.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Indeed, a good look at the consumer part of the system ought to be all that is needed to determine not only representational status but representational content. We argue this as follows. First, the part of the system which consumes representations must understand the representations proffered to it. Suppose, for example, that there were abundant "natural information" (in Dretske's[11] sense) contained in numerous natural signs all present in a certain state of a system. This information could still not serve the system *as* information, unless the signs were understood by the system, and, furthermore, understood as bearers of whatever specific information they, in fact, do bear. (Contrast Fodor's notion that something could function like a representation without functioning like a representation of anything in particular.) So there must be something about the consumer that *constitutes* its taking the signs to indicate, say, p, q, and r rather than s, t, and u. But, if we know what constitutes the consumer's *taking* a sign to indicate p, what q, what r, etc., then, granted that the consumer's takings are in some way systematically derived from the structures of the signs so taken, we can construct a semantics for the consumer's language. Anything the signs may indicate qua natural signs or natural information carriers then drops out as entirely irrelevant; the representation-producing side of the system had better pay undivided attention to the language of its consumer. The sign producer's function will be to produce signs that are true *as the consumer reads the language.*

The problem for the naturalist bent on describing intentionality, then, does not concern representation production at all. Although a representation always is something that is produced by a system whose proper function is to make that representation correspond by rule to the world, what the rule of correspondence is, what gives definition to this function, is determined entirely by the representation's consumers.

For a system to use an inner item as a representation, I propose, is for the following two conditions to be met. First, unless the representation accords, *so* (by a certain rule), with a represented, the consumer's normal use of, or response to, the representation will not be able to fulfill all of the consumer's proper functions in so responding—not, at least, in accordance with a normal explanation. (Of course, it might still fulfill these functions by freak accident, but not in the historically normal way.) Putting this more formally, that the representation and the represented accord with one another, so, is a

[11] *Knowledge and the Flow of Information* (Cambridge: MIT, 1981).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

normal condition for proper functioning of the consumer device as it reacts to the representation.[12] Note that the proposal is not that the content of the representation rests on the function of the representation or of the consumer, on what these do. The idea is not that there is such a thing as behaving like a representation of *X* or as being treated like a representation of *X*. The content hangs only on there being a certain condition that would be *normal* for performance of the consumer's functions—namely, that a certain correspondence relation hold between sign and world—whatever those functions may happen to be. For example, suppose the semantic rules for my belief representations are determined by the fact that belief tokens in me will aid the devices that use them to perform certain of their tasks in accordance with a normal explanation for success only under the condition that the forms or "shapes" of these belief tokens correspond, in accordance with said rules, to conditions in the world. Just what these user tasks are need not be mentioned.[13]

Second, represented conditions are conditions that vary, depending on the *form* of the representation, in accordance with specifiable correspondence rules that give the semantics for the relevant *system* of representation. More precisely, representations always admit of significant transformations (in the mathematical sense), which accord with transformations of their corresponding representeds, thus displaying significant articulation into variant and invariant aspects. If an item considered as compounded of certain variant and invariant aspects can be said to be "composed" of these, then we can also say that every representation is, as such, a member of a representational system having a "compositional semantics." For it is not that the represented condition is itself a normal condition for proper operation of the representation consumer. A certain correspondence between the representation and the world is what is normal. Coordinately, there is no such thing as a representation consumer

[12] Strictly, this normal condition must derive from a "most proximate normal explanation" of the consumer's proper functioning. See LTOBC, ch. 6, where a more precise account of what I am here calling "representations" is given under the heading "intentional icons."

[13] In this particular case, one task is, surely, contributing, in conformity with certain general principles or rules, to practical inference processes, hence to the fulfillment of current desires. So, if you like, all beliefs have the *same* proper function. Or, since the rules or principles that govern practical inference dictate that a belief's "shape" determines what other inner representations it may properly be combined with to form what products, we could say that each belief has a *different* range of proper functions. Take your pick. Cf. Fodor, "Information and Representation," in Philip Hanson, ed., *Information, Language, and Cognition* (Vancouver: British Columbia UP, 1989); and NETC.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

that can understand only one representation. There are always other representations, composed other ways, saying other things, which it could have understood as well, in accordance with the same principles of operation. A couple of very elementary examples should make this clear.[14]

First, consider beavers, who splash the water smartly with their tails to signal danger. This instinctive behavior has the function of causing other beavers to take cover. The splash means danger, because only when it corresponds to danger does the instinctive response to the splash on the part of the interpreter beavers, the consumers, serve a purpose. If there is no danger present, the interpreter beavers interrupt their activities uselessly. Hence, that the splash corresponds to danger is a normal condition for proper functioning of the interpreter beavers' instinctive reaction to the splash. (It does not follow, of course, that it is a usual condition. Beavers being skittish, most beaver splashes possibly occur in response to things not in fact endangering the beaver.) In the beaver splash semantic system, the time and place of the splash varies with, "corresponds to," the time and place of danger. The representation is articulate: properly speaking, it is not a splash but a splash-at-a-time-and-a-place. Other representations in the same system, splashes at other times and places, indicate other danger locations.

Second, consider honey bees, which perform "dances" to indicate the location of sources of nectar they have discovered. Variations in the tempo of the dance and in the angle of its long axis vary with the distance and direction of the nectar. The interpreter mechanisms in the watching bees—these are the representation consumers—will not perform their full proper functions of aiding the process of nectar collection in accordance with a normal explanation, unless the location of nectar corresponds correctly to the dance. So, the dances are representations of the location of nectar. The full representation here is a dance-at-a-time-in-a-place-at-a-tempo-with-an-orientation.

Notice that, on this account, it is not necessary to assume that most representations are true. Many biological devices perform their proper functions not on the average, but just often enough. The protective coloring of the juveniles of many animal species, for example, is an adaptation passed on because *occasionally* it prevents a

[14] These examples are of representations that are not "inner" but out in the open. As in the case of inner representations, however, they are produced and consumed by mechanisms designed to cooperate with one another; each such representation stands intermediate between two parts of a single biological system.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

juvenile from being eaten, though most of the juveniles of these species get eaten anyway. Similarly, it is conceivable that the devices that fix human beliefs fix true ones not on the average, but just often enough. If the true beliefs are functional and the false beliefs are, for the most part, no worse than having an empty mind, then even very fallible belief-fixing devices might be better than no belief–fixing devices at all. These devices might even be, in a sense, "designed to deliver some falsehoods." Perhaps, given the difficulty of designing highly accurate belief-fixing mechanisms, it is actually advantageous to fix too many beliefs, letting some of these be false, rather than fix too few beliefs. Coordinately, perhaps our belief-consuming mechanisms are carefully designed to tolerate a large proportion of false beliefs. It would not follow, of course, that the belief consumers are designed to *use* false beliefs, certainly not that false beliefs can serve all of the functions that true ones can. Indeed, surely if none of the mechanisms that used beliefs ever cared at all how or whether these beliefs corresponded to anything in the world, beliefs would not be functioning as representations, but in some other capacity.

Shifting our focus from producing devices to consuming devices in our search for naturalized semantic content is important. But the shift from the *function* of consumers to *normal conditions* for proper operation is equally important. Matthen, for example, characterizes what he calls a "quasi-perceptual state" as, roughly, one whose job is to cause the system to do what it must do to perform its function, given that it is in certain circumstances, which are what it represents. Matthen is thus looking pretty squarely at the representation consumers, but at what it is the representation's job to get these consumers to do, rather than at normal conditions for their proper operation. As a result, Matthen now retreats. The description he has given of quasi-perceptual states, he says, cannot cover "real perception such as that which we humans experience. Quite simply, there is no such thing as *the* proper response, or even a range of functionally appropriate responses, to what perception tells us" (*op. cit.*, p. 20).[15] On the contrary, representational content rests not on univocity of consumer function but on sameness of normal conditions for those functions. The same percept of the world may be used to guide any of very many and diverse activities, practical or theoretical. What stays the same is that the percept must correspond to environmental configurations in accordance with the same correspondence rules

[15] Dretske (in "Misrepresentation," p. 28) and David Papineau [in *Reality and Representation* (New York: Blackwell, 1987), p. 67ff] have similar concerns.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

for each of these activities. For example, if the position of the chair in the room does not correspond, so, to my visual representation of its position, that will hinder me equally in my attempts to avoid the chair when passing through the room, to move the chair, to sit in it, to remove the cat from it, to make judgments about it, etc. Similarly, my belief that New York is large may be turned to any of diverse purposes, but those which require it to be a *representation* require also that New York indeed be large if these purposes are to succeed in accordance with a normal explanation for functioning of my cognitive systems.

III.

We have just cleanly bypassed the whole genre of causal/informational accounts of mental content. To illustrate this, we consider an example of Dretske's. Dretske tells of a certain species of northern hemisphere bacteria which orient themselves away from toxic oxygen-rich surface water by attending to their magnetosomes, tiny inner magnets, which pull toward the magnetic north pole, hence pull down (*ibid.*). (Southern hemisphere bacteria have their magnetosomes reversed.) The function of the magnetosome thus appears to be to effect that the bacterium moves into oxygen-free water. Correlatively, intuition tells us that what the pull of the magnetosome represents is the whereabouts of oxygen-free water. The direction of oxygen-free water is not, however, a factor in *causing* the direction of pull of the magnetosome. And the most reliable natural information that the magnetosome carries is surely not about oxygen-free water but about distal and proximal causes of the pull, about the direction of geomagnetic or better, just plain magnetic, north. One can, after all, easily deflect the magnetosome away from the direction of lesser oxygen merely by holding a bar magnet overhead. Moreover, it is surely a function of the magnetosome to respond to that magnetic field, that is part of its normal mechanism of operation, whereas responding to oxygen density is not. None of this makes any sense on a causal or informational approach.

But on the biosemantic theory it does make sense. What the magnetosome represents is only what its *consumers* require that it correspond to in order to perform *their* tasks. Ignore, then, how the representation (a pull-in-a–direction-at-a-time) is normally produced. Concentrate, instead, on how the systems that react to the representation work, on what these systems need in order to do their job. What they need is only that the pull be in the direction of oxygen-free water at the time. For example, they care not at all how it came about that the pull is in that direction; the magnetosome that

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

points toward oxygen-free water quite by accident and not in accordance with any normal explanation will do just as well as one that points that way for the normal reasons. (As Socrates concedes in the *Meno,* true opinion is just as good as knowledge so long as it stays put.) What the magnetosome represents then is univocal; it represents only the direction of oxygen-free water. For that is the only thing that corresponds (by a compositional rule) to it, the absence of which would matter—the absence of which would disrupt the function of those mechanisms which rely on the magnetosome for guidance.

It is worth noting that what is represented by the magnetosome is not proximal but distal; no proximal stimulus is represented at all. Nor, of course, does the bacterium perform an inference from the existence of the proximal stimulus (the magnetic field) to the existence of the represented. These are good results for a theory of content to have, for otherwise one needs to introduce a derivative theory of content for mental representations that do not refer, say, to sensory stimulations, and also a foundationalist account of belief fixation. Note also that, on the present view, representations manufactured in identical ways by different species of animal might have different contents. Thus, a certain kind of small swift image on the toad's retina, manufactured by his eye lens, represents a bug, for that is what it must correspond to if the reflex it (invariably) triggers is to perform its proper functions normally, while exactly the same kind of small swift image on the retina of a male hoverfly, manufactured, let us suppose, by a nearly identical lens, represents a passing female hoverfly, for that is what it must correspond to if the female-chasing reflex it (invariably) triggers is to perform its proper functions normally. Turning the coin over, representations with the same content may be normally manufactured in a diversity of ways, even in the same species. How many different ways do you have, for example, of telling a lemon or your spouse? Nor is it necessary that any of the ways one has of manufacturing a given representation be especially reliable ways in order for the representation to have determinate content. These various results cut the biosemantic approach off from all varieties of verificationism and foundationalism with a clean, sharp knife.

IV.

But perhaps it will be thought that belief fixation and consumption are not biologically proper activities, hence that there are no normal explanations, in our defined sense, for proper performances of human beliefs. Unlike bee dances, which are all variations on the

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

same simple theme, beliefs in dinosaurs, in quarks, and in the instability of the dollar are recent, novel, and innumerably diverse, as are their possible uses. How could there be anything *biologically* normal or abnormal about the details of the consumption of such beliefs?

But what an organism does in accordance with evolutionary design can be very novel and surprising, for the more complex of nature's creatures are designed to learn. Unlike evolutionary adaptation, learning is not accomplished by *random* generate-and-test procedures. Even when learning involves trial and error (probably the exception rather than the rule), there are principles in accordance with which responses are selected by the system to try, and there are specific principles of generalization and discrimination, etc., which have been built into the system by natural selection. How these principles normally work, that is, how they work given normal (i.e., historically optimal) environments, to produce changes in the learner's nervous system which will effect the furthering of ends of the system has, of course, an explanation—the normal explanation for proper performance of the learning mechanism and of the states of the nervous system it produces.

Using a worn-out comparison, there is an infinity of functions which a modern computer mainframe is capable of performing, depending upon its input and on the program it is running. Each of these things it can do, so long as it is not damaged or broken, "in accordance with design," and to each of these capacities there corresponds an explanation of how it would be activated or fulfilled normally. The human's mainframe takes, roughly, stimulations of the afferent nerves as input, both to program and to run it.[16] It responds, in part, by developing concepts, by acquiring beliefs and desires in accordance with these concepts, by engaging in practical inference leading ultimately to action. Each of these activities may, of course, involve circumscribed sorts of trial and error learning. When conditions are optimal, all this aids survival and proliferation in accordance with an historically normal explanation—one of high generality, of course. When conditions are not optimal, it may yield, among other things, empty or confused concepts, biologically useless desires, and false beliefs. But, even when the desires are biologically useless (though probably not when the concepts expressed in them are empty or confused), there are still biologically normal ways for

[16] This is a broad metaphor. I am not advocating computationalism.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

them to get fulfilled, the most obvious of which require reliance on true beliefs.[17]

Yet how do we know that our contemporary ways of forming concepts, desires, and beliefs do occur in accordance with evolutionary design? Fodor, for example, is ready with the labels "pop Darwinism" and "naive adaptationism" to abuse anyone who supposes that our cognitive systems were actually selected for their belief and desire using capacities.[18] Clearly, to believe that every structure must have a function would be naive. Nor is it wise uncritically to adopt hypotheses about the functions of structures when these functions are obscure. It does not follow that we should balk at the sort of adaptationist who, having found a highly complex structure that quite evidently is currently and effectively performing a highly complex and obviously indispensable function, then concludes, *ceteris paribus,* that this function has been the most recent historical task stabilizing the structure. To suspect that the brain has not been preserved for thinking with or that the eye has not been preserved for seeing with—to suspect this, moreover, in the absence of any alternative hypotheses about causes of the stability of these structures—would be totally irresponsible. Consider: nearly every human behavior is bound up with intentional action. Are we really to suppose that the degree to which our behaviors help to fulfill intentions, and the degree to which intentions result from logically related desires plus beliefs, is a sheer coincidence—that these patterns are irrelevant to survival and proliferation or, though relevant, have had

[17] A word of caution. The normal conditions for a desire's fulfillment are not necessarily fulfillable conditions. In general, normal conditions for fulfillment of a function are not quite the same as conditions which, when you add them and stir, always effect proper function, because they may well be impossible conditions. For example, Fodor, in "Information and Representation" and NETC, has questioned me about the normal conditions under which his desire that it should rain tomorrow will perform its proper function of *getting* it to rain. Now, the biologically normal way for such a desire to be fulfilled is exactly the same as for any other desire: one has or acquires true beliefs about how to effect the fulfillment of the desire and acts on them. Biologically normal conditions for fulfillment of the desire for rain thus include the condition that one has true beliefs about how to make it rain. Clearly this is an example in which the biological norm fails to accord with the statistical norm: most desires about the weather are fulfilled, if at all, by biological accident. It may even be that the laws of nature, coupled with my situation, prohibit my having any true beliefs about how to make it rain; the needed general condition cannot be realized in the particular case. Similarly, normal conditions for proper function of beliefs in impossible things are, of course, impossible conditions: these beliefs are such that they cannot correspond, in accordance with the rules of mentalese, to conditions in the world.

[18] *Psychosemantics* and NETC.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

no stabilizing effect on the gene pool? But the only alternative to biological design, in our sense of 'design', is sheer coincidence, freak accident—unless there is a ghost running the machine!

Indeed, it is reasonable to suppose that the brain structures we have recently been using in developing space technology and elementary particle physics have been operating in accordance with the very same general principles as when prehistoric man used them for more primitive ventures. They are no more performing new and different functions or operating in accordance with new and different principles nowadays than are the eyes when what they see is television screens and space shuttles. Compare: the wheel was invented for the purpose of rolling ox carts, and did not come into its own (pulleys, gears, etc.) for several thousand years thereafter, during the industrial revolution. Similarly, it is reasonable that the cognitive structures with which man is endowed were originally nature's solution to some very simple demands made by man's evolutionary niche. But the solution nature stumbled on was elegant, supremely general, and powerful, indeed; I believe it was a solution that cut to the very bone of the ontological structure of the world. That solution involved the introduction of representations, inner and/or outer, having a subject/predicate structure, and subject to a negation transformation. (Why I believe that that particular development was so radical and so powerful has been explained in depth in LTOBC, chapters 14–19. But see also section V.6 below.)

V.

One last worry about our sort of position is voiced by Daniel Dennett[19] and discussed at length by Fodor.[20] Is it really plausible that bacteria and paramecia, or even birds and bees, have inner representations in the same sense that we do? Am I really prepared to say that these creatures, too, have mental states, that they think? I am not prepared to say that. On the contrary, the representations that they have must differ from human beliefs in at least six very fundamental ways.[21]

(1) *Self-representing Elements.* The representations that the magnetosome produces have three significant variables, each of

[19] *Brainstorms* (Montgomery, VT: Bradford Books, 1978).

[20] "Why Paramecia Don't Have Mental Representations," in P. French, T. Uehling Jr., and H. Wettstein, eds., *Midwest Studies in Philosophy,* X (Minneapolis: Minnesota UP, 1986), pp. 3–23.

[21] Accordingly, in LTOBC I did not call these primitive forms "representations" but "intentional signals" and, for items like bee dances, "intentional icons," reserving the term 'representation' for those icons, the representational values of which must be identified if their consumers are to function properly—see V.5 below.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

which refers to itself. The time of the pull refers to the time of the oxygen-free water, the locale of the pull refers to the locale of the oxygen-free water, and the direction of pull refers to the direction of oxygen-free water. The beaver's splash has two self-referring variables: a splash at a certain time and place indicates that there is danger at that same time and place. (There is nothing necessary about this. It might have meant that there would be danger at the nearest beaver dam in five minutes.) Compare the standard color coding on the outsides of colored markers: each color stands for itself. True, it may be that sophisticated indexical representations such as percepts and indexical beliefs also have their time or place or both as significant self-representing elements, but they also have other significant variables that are not self-representing. The magnetosome does not.

(2) *Storing Representations.* Any representation the time or place of which is a significant variable obviously cannot be stored away, carried about with the organism for use on future occasions. Most beliefs are representations that can be stored away. Clearly this is an important difference.

(3) *Indicative and Imperative Representations.* The theory I have sketched here of the content of inner representations applies only to indicative representations, representations which are supposed to be determined by the facts, which tell what is the case. It does not apply to imperative representations, representations which are supposed to determine the facts, which tell the interpreter what to do. Neither do causal-informational theories of content apply to the contents of imperative representations. True, some philosophers seem to have assumed that having defined the content of various mental symbols by reference to what causes them to enter the "belief box," then when one finds these same symbols in, say, the "desire box" or the "intention box," one already knows what they mean. But how do we know that the desire box or the intention box use the same representational system as the belief box? To answer that question we would have to know what constitutes a desire box's or an intention box's using one representational system rather than another which, turned around, is the very question at issue. In LTOBC and "Thoughts Without Laws; Cognitive Science With Content,"[22] I developed a parallel theory of the content of imperative representations. Very roughly, one of the proper functions of the consumer system for an imperative representation is to help *produce* a corre-

[22] *The Philosophical Review,* XLV, 1 (1986):47–80.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

spondence between the representation and the world. (Of course, this proper function often is not performed.) I also argued that desires and intentions are imperative representations.

Consider, then, the beaver's splash. It tells that there is danger here now. Or why not say, instead, that it tells other nearby beavers what to do now, namely, to seek cover? Consider the magnetosome. It tells which is the direction of oxygen-free water. Or why not say, instead, that it tells the bacterium which way to go? Simple animal signals are invariably both indicative and imperative. Even the dance of the honey bee, which is certainly no simple signal, is both indicative and imperative. It tells the worker bees where the nectar is; equally, it tells them where to go. The step from these primitive representations to human beliefs is an enormous one, for it involves the separation of indicative from imperative functions of the representational system. Representations that are undifferentiated between indicative and imperative connect states of affairs directly to actions, to specific things to be done in the face of those states of affairs. Human beliefs are not tied directly to actions. Unless combined with appropriate desires, human beliefs are impotent. And human desires are equally impotent unless combined with suitable beliefs.[23]

(4) *Inference.* As indicative and imperative functions are separated in the central inner representational systems of humans, they need to be reintegrated. Thus, humans engage in practical inference, combining beliefs and desires in novel ways to yield first intentions and then action. Humans also combine beliefs with beliefs to yield new beliefs. Surely nothing remotely like this takes place inside the bacterium.

(5) *Acts of Identifying.* Mediate inferences always turn on something like a middle term, which must have the same representational value in both premises for the inference to go through. Indeed, the representation consumers in us perform many functions that require them to use two or more overlapping representations together, and in such a manner that, unless the represenTeds corresponding to these indeed have a common element, these functions will not be properly performed. Put informally, the consumer device *takes* these represented elements to be the same, thus identifying their representational values. Suppose, for example, that you intend to speak to

[23] Possibly human intentions are in both indicative and imperative mood, however, functioning simultaneously to represent settled facts about one's future and to direct one's action.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Henry about something. In order to carry out this intention you must, when the time comes, be able to recognize Henry in perception as the person to whom you intend to speak. You must identify Henry as represented in perception with Henry as represented in your intention. Activities that involve the coordinated use of representations from different sensory modalities, as in the case of eye-hand coordination, visual-tactile coordination, also require that certain objects, contours, places, or directions, etc., be identified as the same through the two modalities. Now, the foundation upon which modern representational theories of thought are built depends upon a denial that what is thought of is ever placed before a naked mind. Clearly, we can never know what an inner representation represents by a direct comparison of representation to represented. Rather, acts of identifying are our ways of "knowing what our representations represent." The bacterium is quite incapable of knowing, in this sense, what its representations are about. This might be a reason to say that it does not understand its own representations, not really.

(6) *Negation and Propositional Content.* The representational system to which the magnetosome pull belongs does not contain negation. Indeed, it does not even contain contrary representations, for the magnetosome cannot pull in two directions at once. Similarly, if two beavers splash at different times or places, or if two bees dance different dances at the same time, it may well be that there is indeed beaver danger two times or two places and that there is indeed nectar in two different locations.[24] Without contrariety, no conflict, of course and more specifically, no contradiction. If the law of noncontradiction plays as significant a role in the development of human concepts and knowledge as has traditionally been supposed, this is a large difference between us and the bacterium indeed.[25] In LTOBC, I argued that negation, hence explicit contradiction, is dependent upon subject-predicate, that is, propositional, structure and vice versa. Thus, representations that are simpler also do not have propositional content.

In sum, these six differences between our representations and those of the bacterium, or Fodor's paramecia, ought to be enough amply to secure our superiority, to make us feel comfortably more endowed with mind.

RUTH GARRETT MILLIKAN
University of Connecticut/Storrs

[24] On the other hand, the bees cannot go two places at once.
[25] In LTOBC, I defend the position that the law of noncontradiction plays a crucial role in allowing us to develop new methods of mapping the world with representations.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

METAPSYCHOLOGICAL RELATIVISM AND THE SELF*

Recent discussion of the problem of personal identity has focused on the distinction between two kinds of theories. Derek Parfit[1] characterizes *reductionism* as holding

1. that the fact of a person's identity over time just consists in the holding of certain more particular facts, and
2. that these facts can be described without either presupposing the identity of this person, or explicitly claiming that the experiences of this person's life are had by this person, or even explicitly claiming that this person exists. These facts can be described in an *impersonal* way.
3. A person's existence just consists in the existence of a brain and body, and the occurrence of a series of interrelated physical and mental events (*ibid.*, pp. 210/1).

Let us say that there is a *direct psychological connection* between two person stages X and Y if and only if Y can now remember having an experience X had, or Y acts in a way that realizes an intention of X's, or Y has a belief, desire, or some other psychological feature possessed by X. The *strong connectedness* of two person stages requires that enough such direct psychological connections hold between them. *Psychological continuity* is the holding of overlapping chains of strong connectedness (*ibid.*, pp. 205/6).

Suppose we know what psychological connections and continuities exist between two person stages A and B, as well as what physical relations obtain between them. And suppose we know the same about the relations between the earlier person stage A and any later stage C whose claim to be a stage of the same person as A competes with B's. Then, according to the reductionist view, which Parfit endorses, we know everything relevant to deciding whether A and B are stages of the same person. Parfit, following Hume, sees an analogy between questions of personal identity and questions about the identity of nations over time. Whether England was the same nation after the Norman Conquest is a question that can only be settled by reference to the connections and continuities among people and institutions. In the same way, questions of personal identity must be

* I am grateful for their comments and criticisms to the members of the audience at the Philosophy Department Colloquium at Columbia University where an earlier version of this paper was presented in April, 1985. I have also benefited from the criticisms and suggestions of Paul Boghossian, Mark Johnston, Derek Parfit, Carol Rovane, and Peter Unger.

[1] *Reasons and Persons* (New York: Oxford, 1984).

0022-362X/89/8606/298–323 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

settled solely by reference to the psychological and physical connections and continuities among person stages.

The *nonreductionist,* in contrast, holds that there are facts of personal identity over and above the impersonal facts. In what follows, I shall criticize both nonreductionism and Parfit's reductionist alternative. Although the theoretical options that Parfit provides are the only ones ordinarily offered, I shall argue that there is a more attractive theory than either of those between which we have thus far been forced to choose.

I. NONREDUCTIONISM AND PERSONAL FACTS

The nonreductionist view that there are facts of personal identity in addition to the impersonal facts involves a number of problems. Since I agree with Parfit that this view is untenable, I shall raise what I take to be the most serious objections before discussing reductionism. These objections and Parfit's are largely complementary.

Proponents of nonreductionism face a dilemma. Either the facts to whose existence they are committed are subjective—that is, facts to which the person whose identity is in question has some special form of access—or they are objective. Subjective facts, I shall argue, are not the kinds of facts which could determine the identity of a person over time. On the other hand, objective facts of the kind required for nonreductionism, if they did settle questions of personal identity, would make it completely mysterious why these questions held any significance or interest.

The most convincing exponent (who may or may not have been an advocate) of the subjective version of nonreductionism is Thomas Nagel.[2] Nagel claims that the true nature of the problem of personal identity is obscured by discussions that focus on continuities and similarities, whether these are physical, mental, causal, or emotional. Given that any such set of conditions is met, there is still the further question whether the same subject or self is preserved under these conditions. The view which Nagel's remarks suggest is one according to which there are irreducibly subjective facts in addition to the objective facts. Whether sugar tastes like *this* to other people is one such fact, and the question whether the subject of *this* experience will exist in some future situation is another (*ibid.*, pp. 200/1).

There are two characteristics that, for Nagel, connect the problem of personal identity with the problem of the subjective character of experience. Unlike facts about objective entities, facts about the subject or self as it figures in the problem of personal identity and facts about the subjective character of experience are not facts about

[2] *Mortal Questions* (New York: Cambridge, 1979).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

entities in the world. And unlike facts about objective entities, the relevant facts about the subject and about subjective experience are accessible only from the first-person point of view; they are not accessible, even in principle, from any point of view other than that of the subject in question. It is in virtue of these two shared characteristics that the aspects of the subject which are relevant to personal identity, as well as the subjective aspects of experience, fall on the subjective side of Nagel's subjective/objective distinction (*ibid.*, p. 201).

It follows on this view that, although we may be fallible in our judgments about subjective facts, we have what might be called *third-person incorrigibility* where such facts are concerned. That is, we could not be corrected by anyone else about the subjective facts to which we have access. The problem for this view is that our judgments about the facts of personal identity are not third-person incorrigible. Suppose, to take an example of Parfit's, that we create at time *t* a molecule-for-molecule duplicate of Smith in a qualitatively indistinguishable environment. If we leave Smith unchanged, this duplicate will believe, and believe mistakenly, that he is Smith and that he existed long before *t*. And these seem to be straightforwardly mistaken beliefs about the facts of personal identity which could be corrected by any external observer with access to enough of the objective facts.

If such a duplicate would not be making a mistake about a fact of personal identity—a fact of the kind to which the proponent of the subjective view claims we have third-person incorrigible access—it is unclear what would count as such a fact. The only kinds of facts which it is even plausible to claim are third-person incorrigible are facts about the contents of our present states of awareness. Personal identity over time, however, must consist in some further facts about the relations between such states. And these further facts would not be determined by the content of any single state of awareness. Thus, the facts of personal identity which were not fixed by the totality of impersonal facts could not be subjective facts.

Suppose, then, that the facts about personal identity that exist, according to nonreductionism, over and above the impersonal facts are not subjective in Nagel's sense. In this case, it would be difficult to see how such a fact, to which we had no special access from the first-person point of view, could explain the intuitions that motivate nonreductionism. If there is any reason to postulate facts over and above the impersonal facts, it is that we feel we have access to something over and above such facts. The postulation of objective facts to which no means of access—whether subjective or not—has been

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

suggested would seem to raise far more serious problems than any which may have motivated nonreductionism. And given this conspicuous lack of access to the facts of personal identity thus construed, it is hard to see how such facts could make a difference to anything that matters.

II. REDUCTIONISM AND THE INSTABILITY OF OUR INTUITIVE JUDGMENTS

In spite of the difference in their respective views, Parfit adopts one of the basic assumptions of those committed to nonreductionism. The assumption is that the crucial division in discussions of personal identity is between nonreductionists and reductionists. On this assumption, once we abandon the intuitions about subjective facts which underlie nonreductionism, the deep problems should disappear. Nagel says, for example, that the concept of the self is the concept of something independent of objective facts and adds that

> [t]his may be an illusion. It may have no sense to speak of 'the same self as this one' in complete detachment from all external conditions. But it is still the internal idea of the self that gives rise to the problem of personal identity (*ibid.*).

And Parfit, though he is not explicit in making this assumption, suggests in a number of remarks that the most difficult puzzles surrounding personal identity will disappear when we drop our prejudices against reductionism (*op. cit.*, pp. 259–260). It is this shared assumption which I shall show is unjustified.

To suppose that the philosophical problem of personal identity is solely the result of bias in favor of nonreductionism underestimates the tension within reductionism itself. Consider three examples.

> *Example 1: The Memory-switch Case.* Suppose your brain is physically altered in such a way that it comes to contain the information and have the dispositions of another subject's, say Smith's. Smith's brain is similarly altered with the result that it contains the information and has the dispositions yours has now. Thus, the person who wakes up in your (present) body after the operation will seem to remember doing all the things Smith now remembers doing and will have Smith's desires, goals, and intentions. Similarly, the person who wakes up in Smith's body will have your (present) psychology. Which, then, of the two resulting persons is you?
>
> It seems implausible to suppose that, for most people, the continued possession of a particular body could be more significant than the possession of a particular sort of psychological make-up. Not even the most serious accidents in which some person survives undermine our belief that the original person survives—unless, of course, there is damage to the brain, and the psychology of the original person is not preserved.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

But, if one survives the radical alteration of the type of body one has, what significance could the numerical identity of one's body have for one's continued existence? We can, if we like, imagine a case in which the new body would be qualitatively identical to the old one. In such a case, neither the subject nor any observer need notice any difference. And, presumably, it is the person with one's memories and psychological dispositions who would realize one's goals and projects. Thus, the conclusion that one would be the person with one's memories and personality seems the natural one to draw.

Example 2: The Future-pain Case. Imagine that you are given the following news. You are to be tortured twenty-four hours from now. There is no chance of reprieve, but you are offered the following consolations. First, before the torture is administered, you will suffer complete amnesia regarding your present experiences. Second, you will be given a set of delusive memories about your past. Third, you will be given a set of dispositions to behave which are at odds with your present character and which constitute a basic change in your present personality. As a final consolation, the same kinds of psychological changes will be effected in another subject, who, after suffering amnesia, will seem to remember the things you remember now and will exhibit your personality traits.

Could any of these changes relieve the apprehension you feel about the impending torture? The argument that they could not is relatively simple. The first three conditions, though they might add to one's apprehension on other grounds, do not seem to lessen the anxiety one would justifiably feel about the future pain. If, on the other hand, none of the first three conditions is any help, it is hard to see how the experience of someone else could have any bearing on the situation whatsoever.

Example 3: The California-therapy Case. Imagine a therapeutic technique that is claimed to remove all the psychological blocks that prevent one from resolving one's problems and using one's opportunities in the most effective way possible. The therapy is effective whether the blocks involve guilt, fear, depression, or anxiety. The therapy requires a session of eight hours a week, during which one is unconscious, and which can be scheduled during the hours one is normally asleep. Imagine you agree to try the therapy, are given a sedative, and awake refreshed with no memories of the session. Suppose that the results exceed even your most optimistic expectations. Your capacity to deal effectively and confidently with the situations you confront far surpasses that of your earlier self.

Suppose, now, that after months of satisfaction with the therapy, the underlying process is explained. After the sedative is administered, the brain matter that realizes your memories and personality is restructured in order to realize the memories and personality of another subject. The resulting person is then awakened and subjected to hours of an excruciatingly painful procedure without anesthesia—a procedure, however, which leaves no discernible effects after the session. At the end of the

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

session that person is sedated, the original structural features of your brain are restored, and you are awakened.

If you had been satisfied with the therapy before all the facts were disclosed, would it be rational to discontinue it after the disclosure? (Those who are incurably squeamish might be offered a pill that would erase the memories of the explanation of the procedure.) If the decision is made on purely selfish grounds, there are no obvious reasons why you should fear the experience of the procedure, given its complete dissociation from your own life history. Furthermore, there is a serious objection to the idea that one should discontinue the therapy. One's experience, given that one undertakes the therapy, is exactly analogous to the actual experience of one of the personalities of a multiple personality subject. In such a subject, the dissociation between the personalities may have been the only available response to a painful situation that the subject as a whole could not avoid. To suppose that forgetting the details of the therapy and continuing would represent merely a kind of philosophical error would be to ignore how deep the sources of our pretheoretical intuitions supporting the therapy may be.

The most important point that these examples bring out is that rejecting nonreductionism for reductionism brings us no closer to a solution to the puzzles they pose. The problem lies in the instability of our intuitions as we move from one case to the next. That our intuitions are unstable in terms of reductionism follows from the fact that the descriptions of all three examples are compatible with their complete equivalence relative to that view. That is, the descriptions are compatible with the stipulation that the degrees of bodily continuity, psychological continuity, and causal continuity are in each case exactly the same.

This fact is, of course, deliberately obscured in the second example. The physical details of the subjects' transformations are ignored, and the description of the events as things that could happen to you might be regarded as question-begging. (Although it is important to notice that this description is part of an event in the example, not part of the description of the example; the fact that, in the example, other people describe the person who is to suffer as "you" will be important later.) To object on these grounds, however, would be to miss the point. On the view suggested by the first example, what matters to one's identity are one's memories and one's psychological dispositions. On this view, the description of the second example as one in which *you* would be given false memories, *you* would be given a different personality, and *you* would be tortured would not be question-begging, but incoherent. The point of the example, however, is that the description is not incoherent; we have no apparent difficulty in understanding the example described. Furthermore, our

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

reaction to the example completely reverses our earlier response. In this case, it is bodily continuity, not the continuity of memories and psychological dispositions, that seems to matter. Even when it is pointed out that the description of the second case is compatible with the description of the first, we seem reluctant to disavow the second response. What does seem to happen, rather, is that we re-evaluate our reaction to the first case in the light of our recognition that the first case is susceptible to the redescription that the second case provides, i.e., the second reaction undermines the first.

The first two examples, which are due to Bernard Williams,[3] provide his argument for the thesis that bodily continuity is the crucial relation in personal identity. The third example, which is intended to bear the same relation to the second which the second bears to the first, is designed to forestall this conclusion. The third example provides another possible redescription of what is, on the reductionist view, an instance of exactly the same case. By the time the third case is reached, however, I think it would be hard to suppose that the examples support either the view that identity depends on the continuity of memory and psychological dispositions or the view that bodily continuity is crucial. Our intuitions are unstable in a way for which the reductionist seems to lack an account. If we take reductionism seriously, these cases should be unproblematic; at the very worst, they should all arouse in us the same sort of ambivalence. The problem is that they do not.

We can appreciate the seriousness of this problem for Parfit, if we look at a number of other cases in which it arises. Consider the following example proposed by David Wiggins.[4]

> *Example 4: The Double-transplant Case.* Suppose that for some subject the two brain hemispheres are psychologically indistinguishable. And imagine that each one is transplanted in a different body, for which it becomes the sole source of control. In the single-transplant case, in which one hemisphere is transplanted and the other is destroyed, the identity of the original subject seems to be preserved. In the present case, each transplant recipient has the same degree of bodily and psychological continuity with the original person as exists in the case of a single transplant. Consequently, the reductionist has no basis for choosing either as the unique successor to the original subject. Thus, since identity is necessarily one-one, this is evidently not a case in which, for the reductionist, personal identity is preserved.

[3] "The Self and the Future," in his *Problems of the Self* (New York: Cambridge, 1973), pp. 46–52.

[4] *Identity and Spatio-Temporal Continuity* (New York: Blackwell, 1967), p. 53.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

This in itself is not a problem for Parfit. Parfit distinguishes personal identity from *survival,* which he characterizes as the preservation of what matters in the identity relation. For Parfit, what survival consists in is the same kind of psychological continuity which he requires for identity, without the requirement of uniqueness. Unlike identity, then, the relation of survival need not be one-one. And, for Parfit, neither identity nor survival requires that psychological continuity have either its normal cause or even a reliable cause (*op. cit.,* pp. 262/3). Hence in the double-transplant case, one survives as each of the resulting people, and everything that matters is preserved in both.

Parfit's characterization of survival, however, requires some clarification. The preservation of what matters might mean the preservation of what makes life worth living. In this sense, what matters might not be preserved in the existence of some future self in whom one would not hesitate to say one survived. One might, for example, be tortured at some time in the future to such an extent that one would prefer, both before the torture and during it, not to go on living. But the preservation of what matters might have a stronger sense. It might mean that there are future person stages for which one would have (or be justified in having) the sort of special concern that one ordinarily has for one's future self. In this sense, what matters is preserved in the torture case. One would normally make very extreme sacrifices where one's present desires were concerned to alleviate some of the pain of the self to be tortured. It is this stronger sense in which it is plausible to identify survival with the preservation of what matters. It is this sense in which I shall use the expression and in which I shall assume Parfit uses it.

In spite of Parfit's distinction between identity and survival, the double-transplant case seems problematic in a way that the normal case does not. Although he claims that everything that matters is preserved, this conclusion is not clearly supported by our intuitions. This problem becomes acute in the following variation of the double-transplant case.

> *Example 5: The Multiple-fission Case.* We can imagine cases of fission in which one splits like an amoeba into two indistinguishable replicas. Consider, then, a case in which one splits simultaneously into a thousand or ten thousand different replicas. Is it really plausible to suppose that everything that matters is preserved in this case and that one survives as ten thousand different individuals?

Another case, which is often mentioned in the literature, also raises serious problems.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Example 6: The Teletransportation Case. Imagine a machine that records the types, speeds, and directions of all the particles that make up a subject at an initial location. This information is then sent to a terminal location where a distinct set of particles is used to produce an exact replica. At the same time, the particles at the initial location are dispersed, and, it is claimed, the subject travels (discontinuously) between the initial and terminal locations. Since this is a nonbranching case which preserves psychological continuity, Parfit is committed to its being a case of identity. More importantly, on Parfit's view, everything that matters is preserved. Thus, we should have all the same responses to this case which we have to the normal case, and it is not at all clear that we do.

Again, consider a variation that has sometimes been proposed.

Example 7: The Defective-teletransportation Case. Suppose the original particles are not dispersed at the initial location. Then one survives as two people. As far as what matters goes, there is nothing to choose between these two, though one may or may not be identical with the person who remains at the initial location. (Whether one is depends on whether the causal connections between the person at the initial location and the person at the terminal location are enough like those between the person at the initial end before and after the switch is thrown to regard the case as a branching one.) Furthermore, on Parfit's view, it should not matter if there is a thousand-year time lag between one's stepping into the machine and a replica's stepping out. If the replica who steps out does so in virtue of a reliable mechanism, the time lag would involve nothing to which Parfit could object. And, since for Parfit even the reference to a reliable mechanism is unnecessary, a replica created a thousand years later in another part of the universe by a monumentally improbable accident is one in whom one survives.

So far we have been looking at cases in which all of Parfit's conditions for survival are satisfied. In spite of this fact, our responses are neither consistent with our responses to ordinary cases, nor consistent with each other (though these are all cases which on Parfit's view are equivalent). Let us now look briefly at a group of mental states which are not related to our conscious life history by the direct psychological relations that for Parfit matter, but which we treat as though they were. The beliefs and desires of our unconscious subsystems, though they may be causally continuous with our conscious experiences, are not in the relevant sense psychologically continuous with them. We do not, for example, remember such unconscious mental states, nor do they contribute to our actions in the way that our consciously accessible beliefs and desires do. Thus, given Parfit's understanding of what matters, we should dissociate ourselves from the effects of such states on our behavior.

In spite of the lack of psychological continuity, however, we do not consider ourselves free to disown such beliefs, desires, and their

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

effects. These unconscious states are as much a part of the self as the conscious states, and we do ordinarily assume responsibility for their consequences. Thus, not only do we have different reactions to cases that should be equivalent on Parfit's view, but we have similar reactions to cases that should be significantly different.

There is a possible reply that Parfit might find tempting. Parfit might claim that we are responding on a case by case basis to irrelevant features of the examples. Given Parfit's views about which features are relevant, this is what he must say to explain the variety of our responses. The difficulty is that Parfit has no explanation of which irrelevant features elicit our responses and why such responses are elicited by those features. I do not intend to suggest that such an explanation is impossible. I shall suggest, however, that, when we attempt to explain why these responses are forthcoming, we shall find that the explanation undermines the claim that the features to which we are responding are, in fact, irrelevant to identity and survival.

III. NONAUTONOMOUS ACCESS RELATIONS

The problem for Parfit stems from his inability to explain our intuitive responses to the puzzles of personal identity. I shall now outline a different theory in order to account for the diversity of these reactions. But the objections so far have been largely addressed to Parfit's view and not to the reductionist view in general. And, given the range of causal conditions which a proponent of reductionism might require for identity or survival, the emphasis on Parfit might seem unjustified. On Parfit's view, as we have seen, the only condition necessary for survival is that the degree of psychological connectedness and continuity is sufficiently high; it is not necessary that it should be caused in the usual way or even in a way that is reasonably reliable. Such a view is arguably the minimal version of reductionism. Thus, the assumption that the problems for Parfit are problems for reductionism in general may seem unwarranted.

There are two reasons, however, why the emphasis on Parfit's version of reductionism should not be a drawback. First, a number of the cases we have looked at raise problems not only for Parfit, but for reductionists who claim that survival requires psychological connections that hold in virtue of either reliable or even normal causal processes. This is true in the memory-switch case (example 1), the future-pain case (example 2), and the California-therapy case (example 3), because the causal explanation of the psychological connections is the same in all three cases, while our reactions differ. Second, suppose that Parfit's claim that the question of survival depends only on the existence of psychological connections—regardless of whether they are reliably caused—is false. (It is at any rate open to

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

serious question.) None of the examples we have looked at, however, except the version of the teletransportation case (example 6) in which the creation of the replica is accidental, involves a cause that is unreliable. And, although all the examples involve abnormal causes, the claim that the normal cause is necessary seems hard to sustain. If there are no facts about personal identity and survival over and above the facts about psychological connections and their underlying causes, then it is hard to see how the normality of these causes could be what matters. The details of the actual causal processes may be inaccessible and may be quite different from what we now take them to be. Thus, if psychological connectedness, reliably caused, were not enough to ensure survival (i.e., to ensure that what matters to us is preserved), it is difficult to see how adding that the causal processes were normal could possibly help.

This argument raises a serious difficulty. Suppose we abandon, as I have suggested in section I we should, the intuitions supporting nonreductionism. What further facts, besides the facts available on Parfit's version of reductionism (or on one of the alternative versions, such as the one according to which psychological continuity requires that the causal relations be normal) could explain those variations in our responses which raise the problems for Parfit? What further facts are there, in other words, to account for the variety in the data? Once we abandon nonreductionism, the cases that cause problems for Parfit seem to raise difficulties for which none of the resources that remain could be adequate.

This problem is not insurmountable, if we recognize that there is an alternative to all the versions of reductionism we have considered so far. Let us begin by reformulating Parfit's characterization of reductionism in terms of supervenience in one of its senses.[5] Reductionism will then entail that fixing some set of impersonal facts fixes the personal facts as well. Or, if we refer to the facts that do the fixing as the supervenience base, the thesis is that the personal facts have their supervenience base in some set of impersonal facts. This way of reformulating reductionism allows us to distinguish clearly between two distinct components of Parfit's characterization. The first is the claim that the fact of a person's identity over time just consists in a set of impersonal facts. The second is the claim that the facts of a person's existence consist in facts about the existence of a brain and a body together with facts about the occurrence of a series of interrelated physical and mental events. Whereas Parfit's first

[5] For a notion of supervenience that fits Parfit's remarks very closely, see Terence Horgan, "Supervenience and Cosmic Hermeneutics," *Southern Journal of Philosophy*, XXII, Supplement (1983): 19–38.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

claim places no limits on the impersonal facts contained in the supervenience base, the second claim limits the supervenience base to facts about the mental and physical events associated with that person's body. (Or, more accurately, the impersonal facts associated with the body of the person in question and those associated with the bodies of any competitors who may have claims to being that person at other times. I shall assume in what follows that the body includes the brain.) In order to distinguish these two claims, I shall call the first *weak reductionism* and the second *strong reductionism*. (Each thesis has two versions, one of which is a claim about personal identity and one of which is a claim about survival.)

> *Weak Reductionism regarding identity (survival).* The fact as to whether person *B* is the same person, existing at a later time, as person *A* (is a survivor of *A*) supervenes on impersonal facts.
> *Strong Reductionism regarding identity (survival).* The fact as to whether person *B* is the same person, existing at a later time, as person *A* (is a survivor of *A*) supervenes on impersonal facts associated with the bodies of *A* and *B* and the bodies of any competitors of *B* who may have claims to being identical with (a survivor of) *A*.

Parfit and most other proponents of reductionism are committed to strong reductionism regarding both identity and survival. (In fact, as I shall point out later, Parfit is committed to an even stronger claim regarding survival.) I shall argue, however, that, if strong reductionism is true in either of its versions, it is true for reasons that are very different from those normally cited in its support.

The strong reductionist claim that limits the supervenience base to the psychological and physical make-up of all the relevant competitors involves the following (ordinarily implicit) assumption: while such relations of psychological access as one's memory of past actions and one's intentions regarding future actions matter for identity and survival, it is only such internal access relations which do. Suppose the assumption turns out to be wrong. In such a case, the following strategy would be possible. We could explain the differences in our responses to cases equivalent on Parfit's view by reference to differences in the relations between person stages which are not found among the internal access relations. On the view I shall suggest, the relations between person stages relevant to survival would include both internal and external relations. Thus, it is the assumption that all the relevant relations are internal which I shall now call into question.

The notion of an internal access relation requires some explanation. Let us say that a mode of access to past events or actions or a

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

mode of access to (including control of) present or future events or actions is *autonomous* if and only if it is possessed by all possible functionally normal adults. Thus, one's access via memory to the event of one's falling from a bicycle at the age of six is via an autonomous mode of access, because access of this type is shared by all actual and possible functionally normal adults. Imagine, however, a subject with an apparent memory of such an actual incident which has been produced on the basis of otherwise inaccessible memory traces by a neurophysiologist aided by a computer. This mode of access is not shared by all functionally normal adults, but only by those living in societies that have the technological capacity (and have made the necessary social arrangements) to make such access possible. Hence, this mode of access is *non*autonomous. Similar remarks apply to one's ability to form an intention on which one's future self will act. Formed in the usual way, this intention provides an autonomous source of access to one's future self; implanted directly in that future self by a hypnotist acting on one's present orders, the intention provides a form of access which is nonautonomous. It is this notion of autonomy which I shall use to explain the sense in which both the proponents of nonreductionism and the proponents of strong reductionism assume that the access relations relevant to identity and survival are internal. Both groups assume that the relevant relations are autonomous in the sense defined.

The solution to the problems for Parfit turns on the existence of modes of access which are nonautonomous. The claim will be that reference to differences in nonautonomous access relations can explain differences in our responses to examples that are equivalent where autonomous access relations are concerned. The examples of nonautonomous access relations involving the implantation of memories and intentions by hypnotists and neurophysiologists provide almost no clue, however, as to how such relations might figure in the examples that cause Parfit difficulties. The only kinds of access ordinarily discussed in connection with such examples are memories and intentions, and, as we have seen, these are normally autonomous. Thus, we need a general characterization of access which has the following consequence: a significant number of our most important sources of access to our future selves—both in the problem cases for Parfit and in the normal cases—are nonautonomous. It may help, then, if we consider what access amounts to in contexts unrelated to personal identity. Suppose we think of access in the context of governmental policy. The two most basic components of this kind of access are knowledge and control. Someone who has very little notion of the workings of government and the kinds of governmental

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

actions which are feasible will have little meaningful access to public policy. Similarly, someone who understands the general workings of government but lacks any knowledge of the constraints imposed by the actual circumstances will lack significant access to governmental decisions. Even someone well-informed, however, may lack access in the relevant sense, if he or she cannot exercise influence over those in power. Such knowledge could not be translated into the power to have his or her desires and intentions determine policy. In the remainder of this section, I shall use this general conception of access to examine a variety of forms—both autonomous and nonautonomous—which access relations may take in the kinds of cases which cause problems for Parfit.

Consider again the teletransportation case (example 6). In certain possible worlds, one has what might be called teletransporter extensions: that is, persons whose presence at the terminal end of a teletransporter has an appropriate explanation in terms of one's own presence at the initial end. In all those worlds in which one has a teletransporter extension, there are certain access relations between oneself and one's extension in virtue of the normal functional make-up one has in common with one's extension. One has access to (and in particular can exercise control over) a teletransporter extension by forming intentions to perform certain actions after the extension comes into existence. One can also exercise control by cultivating long-term dispositions to behave in certain ways. The teletransporter extension, of course, has access to one via memory. Since these relations hold for all functionally normal subjects and their extensions, they are autonomous.

Contrast this with the kinds of access relations which exist between oneself and one's descendants. If one is determined that a child, grandchild, or more distant descendant should perform a certain action, one can invoke reasons or engage in persuasion or bribery, indulge in intimidation or brainwashing, issue orders or threats, or rely on respect or the force of habit. Each of these strategies depends on the specific features of particular societies, and none can be effected merely in virtue of one's normal functional make-up. Hence, these strategies for exercising control yield sources of access which are *non*autonomous. And most of these sources of nonautonomous access which are available in the case of one's descendants will also be available, with minor variations, in the case of one's teletransporter extensions. Although many of these strategies would be difficult to effect directly where a teletransporter extension is concerned, most would be possible to arrange with the help of other people. Thus, the teletransportation case (unlike the case involving

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

one's descendants) exhibits a variety of both autonomous and nonautonomous relations.

The double-transplant case (example 4) provides a second example of the significance of the distinction between autonomous and nonautonomous relations. For those who accept strong reductionism regarding survival, it must be supposed that the claim to be someone in whom we survive is equally strong, whether it is made by one of the two recipients of one of our brain hemispheres or by the person who remains when we have one hemisphere removed and destroyed. But such an analysis is not plausible, except on the assumption that only autonomous access relations are relevant to survival. This is because, while each of the two resulting persons in the double-transplant case has the same autonomous relations to us as the single resulting person in the hemisphere-removal case, the nonautonomous relations may differ substantially.

To see how the single-transplant case and the double-transplant case might differ with respect to nonautonomous access relations, consider a third example in which the departure from the single-transplant case is more extreme. In the multiple-fission case (example 5) in which one splits into a thousand or ten thousand replicas (as in the single- and double-transplant cases), all of our autonomous access relations are preserved. Many of the external sources of access which one would normally have to a single future extension would, however, be absent in such a case of massive replication. An important source of knowledge about our future behavior stems from our knowledge of our future environment—the people to whom we have significant relations, the roles we play, the problems they present, and the opportunities they offer. The difficulty in the multiple-fission case is that an environment so defined could never absorb the number of replicas in question. People of significance to us could not maintain personal relations with each of our replicas. Nor could *we* maintain the pattern of concern for this many future extensions which characterizes our relations to our future selves. Moreover, this number of replicas would be unable to share the roles we play or the opportunities we anticipate. Hence, our ability to predict our future actions and experiences on the basis of our knowledge of our environment would be lost—as would our ability to control such actions and experiences on the basis of our present control of the environment. In the double-transplant case, the loss of nonautonomous access through duplication might or might not be of crucial significance. But, in the case of multiple fission, the replication involved would destroy the major sources of our nonautonomous access to our future selves.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

IV. THE INTRINSIC/EXTRINSIC DISTINCTION

The multiple-fission case (example 5) provides a model of the strategy I shall adopt in section V in dealing with the problems raised by strong reductionism. These problems stem from the fact that, if strong reductionism is true, there is no appropriate explanation of our intuitive reactions to a number of puzzle cases. That is, there is no explanation of our reactions by reference to what, if strong reductionism is true, are the relevant factors—the physical and causal continuities and the autonomous access relations. My strategy in section V will be to examine these cases in the light of the nonautonomous access relations to determine whether they might provide a better set of explanations of our intuitions. Proponents of strong reductionism, however, have suggested a number of principles which, if true, would rule out nonautonomous access relations as relevant to identity and survival. In this section, then, I shall consider whether these principles really do preclude an explanation of our intuitions in terms that are nonautonomous.

To see what would be involved in explaining our responses to difficult cases by reference to nonautonomous access relations, consider once again the example of the teletransporter (example 6). Among the possible sources of nonautonomous access of the subject who steps into the machine to the subject who steps out are the following. First, there is the legal title of the subject stepping out to the possessions of the person who stepped in. If legal title is preserved, then the initial subject can control to a large extent the environment of the terminal subject by providing for that subject. Second, there are the legal and social relations of the individual at the terminal location and the persons who figure importantly in the life of the individual at the initial end. Suppose that the terminal individual remains married to the spouse of the initial individual, that he or she has the same family ties, and that he or she is a member of the same groups, a citizen of the same country, and a friend of the same individuals. This will increase not only the information that the initial subject will have about the terminal subject, but again the initial subject's ability to plan and control the terminal subject's future. Third, there is a source of nonautonomous access in the general contractual rights and obligations of the terminal subject. If these remain intact, there is a further source of rational control for the person contemplating the process.

Whether we think intuitively of the person at the terminal location as a copy of the original person (and in one's own case whether one thinks of the person as a copy of oneself) depends, on the view I am suggesting, to a large extent on how we imagine other people re-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

sponding. If the process is described as a means of transportation, then we are thinking of a society in which the subject at the terminal end is taken to be the original and not merely a copy. If personal identity and survival are determined, in part, by external factors of the kind under consideration, then it is not far from the truth to claim that in a society in which teletransportation is interpreted as transportation, it *is* transportation.

Explaining the puzzle cases in terms of the relevance of nonautonomous access, then, has the consequence that the members of one society might survive teletransportation whereas the members of another would not. This claim, however, is open to objections both from those who think that teletransportation must preclude survival and those who think that in such cases survival is guaranteed. It is a common response among philosophers, for example, to suppose that a society that interprets teletransportation as transportation must be mistaken. According to this response, teletransportation is always incompatible with the preservation of personal identity. The response is based on considerations like the following. Suppose the machine that disperses the particles of the person at the initial location fails. In such a case, the person who is still at the initial location exists simultaneously with the person at the terminal location. Since the former, it is alleged, is obviously identical with the original subject, the person at the terminal end must be a copy. It is then concluded that the person at the terminal end must be a copy *even when the machine functions correctly.*

A similar objection involves the fact that, even if the particles of the subject at the initial end are dispersed, the information about the particles might be transmitted to two terminal locations. Two persons would then be created, each with the same relation to the original. Since they could not both be identical with the original and since there are no grounds to choose between them, neither one could be identical with the person at the initial end. And, because in the case where the machine works the relevant relation between the initial subject and the terminal subject is the same as in the case in which it does not, neither relation is one of identity.

These arguments, due in large part to Williams, rely on principles like the following:

> Principle (1): Whether a future person will be me must depend only on the intrinsic features of the relation between us. It cannot depend on what happens to other people.[6]

[6] This principle, which underlies a number of Williams's arguments, is formulated by Parfit, p. 267. For an explicit formulation of Principle (1) by Williams, see

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

This principle counts against our regarding the person who steps out of the teletransporter as identical with the person who stepped in. For even in the case in which the teletransporter works correctly, it could have failed to disperse the particles of the person who stepped in. And, if it had failed, we would, Williams claims, surely count the person remaining at the initial end as identical with the person who stepped in. Thus, suppose that in cases in which the teletransporter worked it preserved identity. Then, whether the person at the initial location when the switch was thrown was identical with the person at the terminal end would depend on whether the person at the initial end continued to exist past the time the switch was thrown. Hence, the assumption that the teletransporter could ever preserve identity violates principle (1). Given his commitment to principle (1) and the fact that he does not distinguish personal identity and survival, I shall assume that Williams is a proponent of the following thesis, in the form which concerns survival as well as the form concerning identity.

> *Very Strong Reductionism regarding identity (survival).* The fact as to whether person *B* is the same person, existing at a later time, as person *A* (is a survivor of *A*) supervenes on impersonal facts associated with the bodies of *A* and *B*. It does not depend on facts about any other people.

Thus, whereas for the strong reductionist identity (survival) supervenes on the autonomous relations between *A*, *B*, and any relevant competitors, for the very strong reductionist like Williams, identity (survival) supervenes on the autonomous relations between *A* and *B* alone.

Whereas Williams would object to the claim that a society or culture might construe the teletransporter correctly as transportation, Parfit would object to the claim that a society might view it correctly as death—at least where this is taken to be incompatible with survival. Parfit is committed to the claim that teletransportation always guarantees survival regardless of how it is construed by those involved. For Parfit, the fact that the particles of the subject at the initial teletransportation location could have failed to disperse, or that more than one duplicate could have been made, is irrelevant to what matters, namely, survival. Parfit holds this even though he holds the analogue of Williams's principle (1) for survival. That is, he holds

> Principle (2): Whether my relation to some future person *P* contains what matters depends only on the intrinsic features of the relation between us. It cannot depend on what happens to other people.

"Cosmic Philosopher," *The New York Review of Books,* XXIX, 2 (February 29, 1982), p. 33. The review is of Robert Nozick, *Philosophical Explanations* (Cambridge: Harvard, 1981).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

This amounts to holding very strong reductionism regarding survival. Parfit can hold both these theses, because he allows that survival need not be one-one. [Where identity is concerned, Parfit rejects very strong reductionism and principle (1) (*op. cit.*, pp. 266–273).] And, given that the teletransporter preserves psychological continuity, Parfit's claim that survival need not be one-one, together with his claim that psychological continuity is what matters regardless of its cause, yield the conclusion that, in cases of teletransportation, survival is always assured.

Like Parfit, I reject principle (1). A case of fission in which one of the duplicates is immediately destroyed might well be a case in which one's identity is preserved. This is possible even though, if that duplicate had lived, one would not have been identical with either duplicate but would (at best) have survived in both. But, unlike Parfit, I also reject principle (2). Not only can one's identity with some future person depend on what happens to other people, but whether that person's existence preserves what matters to one (i.e., whether one survives as that person) can depend on other people as well. As we saw in section III in the multiple-fission case (example 5), whether what matters is preserved may depend on how many future persons are psychologically continuous with one's present self. Too many duplicates may prevent what matters from being preserved, even though the intrinsic relations between one's present self and each of the duplicates do not seem significantly different from what they are in the normal case. This is a case in which the others who make a difference are competitors. Another such case that we have looked at is the defective-teletransporter case (example 7). Consider the person created in a far-flung corner of the universe many years hence. Given that a normal duplicate is also created, one might plausibly feel that the existence of the more remote duplicate matters scarcely at all. In the absence of such a normal duplicate, however, one might invest in the more remote duplicate—who would represent one's only chance of survival—something like the concern one usually reserves for one's ordinary future selves.

In addition to providing the candidates for our future selves, other people provide the means in virtue of which we gain nonautonomous access to such selves and their competitors. If, for example, there were a social convention concerning cases of multiple duplication whereby one of the duplicates inherited the entire network of social relations, one might then regard oneself as surviving in that duplicate and regard the others as largely irrelevant. Such a convention, after all, would prevent just that situation which made it plausible in the case without the convention to suppose that one did not survive: namely, the situation in which none of the thousands of duplicates

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

could inherit any substantial portion of one's significant relations to others.

This rejection of principle (2) is not without precedent. Robert Nozick has claimed that whether what matters to one is preserved in the existence of some future person *P does* depend on who the competitors of *P* are (*op. cit.*, p. 63). In other words, Nozick rejects very strong reductionism regarding survival as well as identity. Even for Nozick, however, the question depends only on the autonomous relations between oneself and the various candidates for being one's future self. Thus, for Nozick, strong reductionism is true regarding both identity and survival. On my view, even this claim is false. Strong reductionism is false for both identity and survival. Whether survival or identity is concerned, weak reductionism provides the best account. Since principle (1) and principle (2) are intuitively plausible, however, I shall provide two more variations on the future-pain case (example 2) to show that the intrinsic/nonintrinsic distinction cannot bear the strain to which Parfit's theory subjects it.

For Parfit, anxiety is not an appropriate response to the future-pain case, in which one's torture will be preceded by a memory and personality switch. Once the amnesia and the personality change have been effected, the psychological continuity and connectedness, which for Parfit are what matters, have been completely destroyed. One has no more reason to fear pain that will be felt by such a subject than to fear pain that will be felt by a perfect stranger.

The normal case, in contrast, is one in which psychological continuity and connectedness remain intact. In the normal case, if one experiences a period of intense pain, one is psychologically connected to, and continuous with, the earlier self that anticipates the pain. Such connections exist both in virtue of one's conscious occurrent memories of one's earlier self and in virtue of one's dispositions to form such conscious memories as occasions arise. The same is true of the continuity between the personality of the earlier self and that of the later self. This continuity is manifest in the similarity of occurrent episodes of thought and behavior, but it is also manifest in the similarity of the dispositions underlying such episodes. One's relation to the future person stages that are to suffer the pain might, however, be somewhat different. Consider the following sequence of cases.[7]

Example 8: The No-occurrent-memories Case. Suppose one knew that one would be tortured at some future date. One might believe with good reason that once the ordeal began, one would be too preoccupied

[7] Parfit considers a number of similar cases but does not seem to recognize their full force. See Parfit, pp. 229–230.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

with the immediate sensations of pain to entertain conscious memories of any prior experience. This thought, if it occurred, would be unlikely to provide any consolation. Whether one consciously reflects on any of one's past experiences is likely to be a function of the pacing of the ordeal and the presence or absence of any reminders of the past. And such factors seem irrelevant to the question of how painful the experience would be. Nor is it plausible to attach to factors such as these the importance that Parfit wants to attach to identity and survival. The same line of reasoning applies to the conscious manifestations of one's characteristic personality traits.

Example 9: The Undetected-memory-switch Case. In this case, imagine suffering the psychological discontinuity described in the future-pain case (example 2). At the flip of a switch, one's memories and personality are replaced by delusory memories and a new personality. The difference in this case is that the switch is thrown at the same time the pain begins and one's conscious experience is as described in the no-occurrent-memories case (example 8). Hence, there will be no conscious manifestation of the psychological difference induced when the switch is flipped. Thus, like the last change, this change from the normal case should produce no rational grounds for consolation in thinking of the future pain.

Example 10: The Sporadic-occurrent-memories Case. In the final variation, imagine either that the switch is flipped shortly before the pain starts, that the ordeal is paced slightly more slowly allowing a few minutes of relief, or that there are more reminders of the past. Although this makes a difference in one's conscious experience, this difference seems too trivial to provide a plausible justification for viewing this prospect more optimistically than either the normal case or the other two variations on the normal case (examples 8 and 9). This case, however, simply *is* the future-pain case (example 2), with regard to which Parfit is committed to claiming that one has no reason to fear the painful experience. The problem is that the importance of the difference between the normal case (in which anxiety about the future pain is justified) and the future-pain case (in which it is not) seems to dissolve on examination. Since this is the most central case of the kind of intrinsic difference which Parfit thinks matters to survival, his assumption that only such intrinsic differences could matter seems unwarranted.

V. NONAUTONOMOUS RELATIONS AND THE EXPLANATION OF OUR INTUITIVE JUDGMENTS

The question that I now want to consider is whether reference to nonautonomous relations can solve the original problems for proponents of the strong reductionist claim that what matters depends only on autonomous features. These problems arose because this claim made it impossible to explain our differing intuitions regarding the three versions of the memory-switch case (examples 1–3). It should be clear by this point that these three examples, though equivalent as regards what Parfit and Williams think matters, are

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

likely to suggest very different situations where nonautonomous access relations are concerned.

The description of the original memory-switch case (example 1) not only leaves open the possibility that the nonautonomous access relations are inherited by the person who inherits the memories, it encourages this assumption. This claim is best explained by reference to what psychologists refer to as default assignments. In understanding the stories that make up these examples, we adopt assumptions that are not strictly given in the stories themselves—assumptions that are subject to revision as new information comes in. And, in the absence of any other cues, it is natural to assume that the body that inherits one's memories and intentions will have roughly the same capacity to act on those intentions and realize one's projects as any of one's normal future selves. This is a reasonable assumption for those whose projects and relations to others are not importantly tied to their particular bodily make-up. Thus, the first example encourages us to assume that the relation between autonomous access and nonautonomous access is fixed. There is nothing to suggest what is nonetheless a real possibility: that autonomous and nonautonomous access may come apart and that someone who inherits all of one's social relations may be more likely to carry on one's relationships and commitments and to realize one's intentions and projects than someone who inherits one's psychological make-up.

In the future-pain case (example 2), by contrast, the normal relation between autonomous access and nonautonomous access cannot be taken as given. If you are told that *you* will experience amnesia and be given delusive memories and new personality traits, there is a clear contrast established between the first-person and third-person perspectives on your identity. From the point of view of the people in charge, and, by implication, the outside world generally, the person with your body remains you in spite of the loss of autonomous access. (This is the significance of the fact that people in the story refer to the person who will suffer as "you.")

The California-therapy case (example 3), like the original memory-switch example, strongly suggests that autonomous access and nonautonomous access are present or absent together. The external perspective that the example suggests is not one in which you are tortured for your own long term good. Rather, it is one in which your body and brain are manipulated in various ways during periods in which you remain unconscious.

Notice that, if we have an explanation of the fact that the changes in the future-pain case provide no consolation, we also have an explanation of the two variants of the future-pain case—the no-occurrent-memories case (example 8) and the undetected-memory-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

switch case (example 9). In both these cases, the intrinsic or autonomous changes are less important than the continuity of the nonautonomous relations. Whether it is we who are suffering depends, in part, on whether those around us could be expected to respond as though it is. In the undetected-memory-switch case and the no-occurrent-memories case, there is no reason at all for an outside observer not to feel that it is the original subject who is being made to suffer; nothing accessible to an ordinary observer in either case would reveal anything unusual. This contrasts with the California-therapy case in which it is relatively easy to dissociate the subject who suffers from the original subject. Not only does the original subject return unharmed with no memories of any painful experience, but our access to the suffering is mediated by the medical personnel for whom the psychological dispositions of the person who suffers seem most significant. Compared with the California-therapy case, the sporadic-occurrent-memories case (example 10) provides fewer barriers to nonautonomous access and less support for dissociation where the subject who suffers is concerned. An emphasis on how little significance the memory switch may have for one's conscious experience under torture suggests an external perspective in which one would not seem significantly different to others. It may also suggest a society in which the facts about one's conscious experiences, rather than the facts about one's psychological dispositions, are highly accessible and salient. In any event, there is no suggestion, as there is in the original memory-switch case and the California-therapy case, of a social practice and a perspective on oneself in which such dispositional changes would justify an attitude of indifference to the prospect of the future pain.

As a final test of the hypothesis that nonautonomous access relations matter, consider two more examples.

Example 11: The Realistic-nightmare Case. Suppose that you are told by a psychologist on the basis of a neurophysiological theory and a brain scan that you have had, and will continue to have, terrifying nightmares, which you never remember. Unlike ordinary dreams, however, these nightmares have no adverse consequences for your waking experiences and, in fact, tend to have positive effects. If Parfit is right, whether you should be anxious ought to depend on the content of the nightmares. If your responses to situations in the dreams are what your actual responses would be, if the characters in the dreams correspond closely to people you actually know, and if the past incidents you seem to remember actually occurred, then the psychologist's news should be serious cause for concern. This is because from the inside the case would be indistinguishable (until you are reawakened) from one in which the events actually occur. If, alternatively, your reactions in the dream are unlike your normal reactions, if the characters in the dream have no real

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

life counterparts, and if you seem to remember a very different past from your actual one, the psychologist's news should occasion no anxiety.

In fact, whether the news that one experiences such nightmares produces, and is justified in producing, anxiety seems completely independent of the details of the nightmares' contents. Furthermore, most of us, if told that we experienced such nightmares, would not feel any great anxiety. This fact is explained by the tenuousness of the nonautonomous relations, in this society, between waking and sleeping subjects. Nothing in the example suggests that outside observers would have any reason to feel anxiety for us. As a result, we feel no anxiety for ourselves. And the tenuousness of such relations is independent of any facts we might be given about the detailed content of the nightmares.

Compare this case with a variation of the California-therapy case.

Example 12: The Cut-rate-therapy Case. In this case, one undergoes the procedure of the California-therapy case without either anesthesia or the benefit of any change in one's memories or personality. After the procedure the memories of the pain are erased, leaving one satisfied with the therapy and eager to continue. In this kind of case, one would be very unlikely to continue the therapy once the facts were explained. But the difference between this case and the realistic-nightmare case is again a difference in the *non*autonomous access relations alone.

Thus, it seems difficult to deny that such relations explain a wide variety of our actual responses.

In spite of the examples, however, the feeling is likely to linger that to take the role of nonautonomous access relations seriously in questions of personal identity is to risk changing the subject. What must ultimately matter, after all, is not the third-person perspective of others, but the first-person perspective of the subject in question. To offer such an objection, however, is to miss an important ambiguity in the notion of the first-person point of view, as well as what is most important in the notion of nonautonomous access. On one possible construal, the first-person point of view is the subjective view: that is, the view defined by the set of facts which are independent of any objective facts whatsoever. The claim that personal facts must be manifest from such a perspective, then, is the nonreductionist claim that personal facts are subjective facts. And this claim, as we have seen, does not lead to a viable theory of personal identity. In another sense, though, the nonautonomous facts about a subject are a crucial part of that subject's first-person point of view. The perspective of others is *a* significant, if not *the* significant, component of our internalized self-image. Other people, real or imagined, and internalized, are an important part of our psychological continuity and of our own access to ourselves. Our internalized image of ourselves *as seen from*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the outside—sleeping tranquilly and awakening refreshed—is part of what robs the realistic-nightmare case of its capacity to produce concern. It is this internalization of others in virtue of which our society's customs and practices are personal facts about what matters.

If nonautonomous access relations are relevant to the explanation of our intuitive judgments about personal identity, then we should reject William's principle (1), Parfit's principle (2), and Nozick's position. The claim that the personal facts about a given subject (including such facts as whether what matters is preserved in the existence of some future person *P*) supervene on facts about the intrinsic physical and psychological relations between the subject and *P* is too strong. And the claim is still too strong, even if we include in the supervenience base all the facts about the intrinsic physical and psychological relations between the subject and all the relevant competitors of *P*. We should accept only some weaker supervenience claim that incorporates facts about the relations of nonautonomous access in the supervenience base.[8] The personal facts about a given subject do not supervene on any set of facts which does not include facts about the attitudes and feelings of others, the conventions and practices of the subject's society, and in some cases even the society's level of technological development.[9]

VI. METAPSYCHOLOGICAL RELATIVISM

These claims lead naturally to a position that I shall call *metapsychological relativism*. The metapsychological facts are those facts about

[8] There is an alternative for dealing with the apparent counterexamples to principles (1) and (2) which does not have the consequence that relations to other people must be included among the facts upon which identity and survival supervene. On this alternative, whether identity or what matters are preserved would depend not on the *actual* relations to others of the subject in question, but on what relations he or she *thinks* obtain and has internalized. The facts about identity and survival would, then, supervene on the physical and psychological facts about the subject, the future person, and any competitors. But this suggestion is not one in which Williams, Parfit, or Nozick could take any comfort. Even if this alternative were to prove attractive, it would be extremely implausible to claim that whether *S* survives as *P* depends only on the intrinsic relations between *S* and *P*, or between *S*, on the one hand, and *P* and *P*'s competitors, on the other. On any ordinary interpretation, the requirement that the relations must be intrinsic would rule out *S*'s beliefs about the social practices in his community, just as it rules out facts about the practices themselves. Thus, if principle (2) and Nozick's position are not regarded as false on this alternative, they will be very different claims from what was apparently intended.

[9] Others who have objected to the narrowness of the set of facts considered relevant to questions of identity and survival by most of the parties to the current debate include: Norman Daniels, "Moral Theory and the Plasticity of Persons," *The Monist*, LXII, 3 (1979): 265–288; Peter McInerny, "My Future, Right or Wrong," *Philosophical Studies*, XLIV, 2 (1983): 235–245; and Amelie Rorty, *The Identities of Persons* (Berkeley: California UP, 1976), pp. 3–4, and "Persons Shmersons," in Christopher Gill, ed., *Persons and Human Beings* (forthcoming).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

personal identity, responsibility, and the unity and character of the self which are presupposed, rather than settled, by empirical psychology. The relativist claim, then, is that there could be two societies, neither of whose members are either misinformed or irrational, but whose social practices settle the significant metapsychological issues in radically different ways.

The teletransporter case (example 6) provides a particularly plausible example. Is there any rational way of choosing between a society in which the social conventions support the use of the teletransporter as transportation and a society in which they do not? It is true, of course, that given some relevant social conventions, it might seem irrational to adopt others. If, for example, the person at the terminal location retains the marital and family ties of the original person, then a convention by which contracts become null and void would seem at least highly unmotivated. The question, however, is whether, between a society in which all the institutions and practices support the use of the teletransporter as transportation and one in which none do, there would be any basis for a rational choice. And it is not apparent what the basis of such a choice could be. There seems, then, to be no neutral way of choosing between the two societies.[10]

Metapsychological relativism is not, of course, the facile conventionalism that claims that one simply decides what is to count as an extension of oneself. One cannot decide that by 'oneself' one means one's body or alternatively one's psychology, except as a trivial stipulative definition. And there is clearly more at stake in the puzzle cases than that. To say that there is more at stake, however, is not to endorse a belief in the kind of transhistorical fact for which antirelativists have searched. We cannot decide as individuals what counts as our own continuation, any more than we decide as individuals what kind of society we shall inhabit. But as we determine collectively the form and shape of our society, we thereby also determine the boundaries of ourselves.

STEPHEN L. WHITE

Tufts University

[10] Sydney Shoemaker makes a similar point in Shoemaker and Richard Swinburne's, *Personal Identity* (Oxford: Blackwell, 1984), p. 132.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

COMMENTS AND CRITICISM

THE MALADROITNESS OF EPISTEMIC TIT FOR TAT*

IT has been well-argued that much of what we are justified in saying that we know we do not and could not know on our own account.[1] I shall not repeat these arguments here, but shall make it a point of departure that a good deal of what we take ourselves as knowing we accept and are right to accept on the say-so of others, on the basis of their expertise or their special position to know. It may be thought that such dependencies are conveniences of laziness and that, strictly speaking, they do not furnish us with knowledge, which might better be thought of as reserved for those cases in which we perform the evidentiary and justificatory work on our own, solo, so to speak. For what I intend to propose in this note, it is not necessary or particularly important that I carry a brief for dependent or vicarious knowledge. It will suffice to construe such dependencies as issuing in rational acceptance; and so I rephrase the Hardwig thesis as affirming that much of what it is rational to accept and to act on comes to us by way of the warrant of authoritative say-so.[2] Resort to authoritative say-so is discretionary in various cases, but given our cognitive finitude it could not be dispensed with in the general case without, among other things, ransacking our prospects for survival. And so it is not just all right, it is essential.

The trust we place in others is both institutional (as in science) and individual ("Is this the way to Central Station?") and, in either case, it throws up an intriguing question: are there ways of illuminating the dynamics of epistemic trust? A provocative and engaging answer to this question has recently been furnished by Michel J. Blais.[3] He asks whether there might be some explanation of why our collective epistemic cooperation is not just unavoidable but also reliable. His answer is that epistemic cooperation is strategically irresistible and that

* Research that gave rise to this work was supported by grants from the research fund of the University of Lethbridge and the Social Sciences and Humanities Research Council of Canada, for which I am grateful. An earlier version was presented to the Institute for Functional Research in Language and Language Use at the University of Amsterdam. I am grateful to M. A. Haft-Van Rees for helpful comments.

[1] John Hardwig, "Epistemic Dependence," this JOURNAL, LXXXII, 7 (July 1985): 335–349.

[2] I shall, however, for convenience and without prejudice, retain the word 'epistemic' as in "epistemic trust" and "epistemic cooperation."

[3] "Epistemic TIT FOR TAT," this JOURNAL, LXXXIV, 8 (August 1987): 363–375.

0022-362X/89/8606/324–331 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

its allure can be understood as an adaptation of a certain kind of solution for iterated Prisoners' Dilemma (PD) games that we might imagine the epistemic (or rational-acceptance) community to be playing in a state of nature. Concerning this imaginative assumption, Blais (1) holds with Robert Axelrod[4] that winning strategies for such games are cooperative rather than defective, and (2) conjectures that our reliance on the say-so of others can be made strategically lucid and defensible as a representation of iterated PD strategic success. My purpose in this note is to show that, although the Axelrod results are convincing, Blais's adaptation of them to our communal epistemic practice should be resisted.

I. THE AXELROD RESULTS

For the sake of brevity, I assume the reader's familiarity with the following things: the basic or classical PD-game; the pay-off matrix for a PD-game; an iteration of PD-games; and the discount parameter for an iteration. I shall also leave unrepeated most of Blais's exposition of Axelrod's findings. These are well–explained by Blais himself, and by Douglas R. Hofstadter[5] and, of course, Axelrod.

In the classical game, in which there is no assumption of rematch, two players move concurrently, but without either knowing the other's move. This happens once, then the game is over. Their safest individual strategy is defection, which minimizes the worst that (in the terms of the game) could happen to either. Jointly resorted to, the safe minimax strategy draws a punishingly low pay-off for the two participants. Yet there is no doubt, punitive pay-off or not, that this is the best joint outcome that the conditions of the game makes possible. So it looks as if PD-environments visit upon even their strategically most perceptive inhabitants the disconsoling prospect of low pay-off. The concrete rewards of safety-from-worse can be pretty dismal, and the evasion of worse rejoicable only in the abstract.

Things change dramatically when players meet again and play a series of PD in which each player has perfect recall of the history of each other's previous moves in the series. Axelrod is able to show that the minimax strategy of the classic game (a limiting case of ALL D, that is, defection at every move) will be displaced in the iterated setup by a cooperative strategy such as TIT FOR TAT. TIT FOR TAT cooperates on the first move, continues to cooperate until the opponent defects, repays that defection once, and then resumes a

[4] Axelrod's results are reported in his important study, *The Evolution of Cooperation* (New York: Basic Books, 1984).

[5] To whom Blais acknowledges an expository debt; see "The Prisoner's Dilemma, Computer Tournaments and the Evolution of Cooperation," in *Metamagical Themas* (New York: Bantam, 1986), pp. 715–734.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

cooperative posture. If the opponent defects again, TIT FOR TAT will make the same retaliatory response as before; and so on for any further defections.

TIT FOR TAT is successful because it possesses the following properties:

1. It is *nice*. That is, it never defects first (or, a subtlety, until quite late in the iteration).
2. It is *retaliatory*. It repays a defection with a defection immediately and always.
3. It is *forgiving*. It repays any given defection once only, and then returns to cooperation.
4. It is *clear*. It is not a hard decision rule for the opponent to figure out.
5. It is *robust*. It does well in a great many environments of multivarious strategic adversity.
6. It is *invasive*. It displaces uncooperative strategies even where these latter are dominant in a playing population.
7. It is *resistant*. It successfully fends off attempts by defective strategies to invade it.

II. BLAIS'S ADAPTATION

Let us see what Blais tries to do with these results. He is careful to say that TIT FOR TAT is an idealized model of human behavior and "cannot be taken very seriously as an exact description of how and why humans cooperate or defect, whether in day-to-day affairs or in the pusuit of science" (*op. cit.*, p. 369). Although TIT FOR TAT is "simplistic," strategies somewhat like it are "the norm rather than the exception, even in the supposedly tough [dog-eat-dog world of the business] marketplace" (*op. cit.*, p. 370). They are also the norm in science, where "cooperation is of the essence" and honesty is the best policy.

For all his care not to overstate the applicability of TIT FOR TAT to ordinary life, it is overstated nonetheless. Blais's basic contention is that, up to a point, epistemic cooperation can usefully be understood as having incorporated a strategy which, though not TIT FOR TAT precisely and not TIT FOR TAT exclusively, is a strategy of like-type which would do well in PD-environments; and this, I think, is untrue. Here are some problems.

A. The *decisional nodes* of PD-play are opaque. Players play their hands (other than at the opening moves) as a partial function of what happened previously, but not as a function of or reaction to the moves they exchange at that node. Here decisions are exchanged blindly.

As to the applicability of cooperative principles (honesty is the best policy) to communication enterprises such as exchanges of informa-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tion, it becomes important to notice that a *communication node* is representable as an ordered pair $\langle A(s),B(s')\rangle$ in which a speech act s of A is responded to by a speech act s' of B,[6] where s' is a partial function f of s, i.e., where $s' = f(s)$, and where the pair $A(s),B(s')$ is ordered by the relation *is a response to*.

Now, if we seek to represent communication as PD-behavior, then a communication node turns out to be a strictly unordered pair $\{A(s),B(s')\}$ in which the elements co-occur in that node independently of each other. Actual informational exchanges are not typically like this, of course.

B. Decision rules that are perfectly intelligible strategies for PD-play are not by and large intelligible for communication exchange. Winning PD-strategies are cautiously retaliatory, that is, they repay defections with measured defections of their own. Even if we put aside the ordered-pair problem for real life communication exchanges, it is hard to see the strategic sense in, for example, repaying a lie with a lie or any other corruption of communication felicity.

The problem is exacerbated by the difficulty of interpreting a one-move communication set up in *classical* (noniterated) PD-terms. For one thing, the pay-off matrix looks odd. If we interpret lying (l) as defection and truth telling (t) as cooperation, then the classical pay-off matrix yields:

1. $A(t),B(l)$: 0 for A and 5 for B
2. $B(t),A(l)$: 0 for B and 5 for A
3. $A(t),B(t)$: 3 for A and 3 for B
4. $A(l),B(l)$: 1 for A and 1 for B

Although we can search for and find situations (see below) in which the matrix gets some realistic grip, it does not in general. In classical game terms, it must be intelligible to reckon that the worst that could happen to me is always my telling the truth while you lie. Without this assumption, the rationality of the classical minimax strategy falls apart.

C. Actual communication exchanges are often *verificationally inefficient*. Even where what I do at the present node is conditioned by what you did at the previous node, it is not typically open to me to know whether your move at the previous node counts as a defection or as a cooperation. A PD-reconstruction is wholly insensitive to this verificational lag.

Here is an imaginary case. A is a senior official of government department X and B of department Y. A and B arrange to exchange confidential departmental information using a "drop" at the Carlton

[6] Not all types of informational transmission are bilateral, of course. See David Lewis's discussion of two-sided signalling problems in *Convention* (Cambridge: Harvard, 1969), ch. 4.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Club. Since all government departments wage ceaseless battles for funds from the treasury, we can imagine a pay-off matrix for the exchange which reflects the fact that the acquisition of true confidential information will strengthen competitive advantage at the treasury, the acquisition of false information will not, and that the worst situation possible is the double cross in which A exchanges for B's misinformation his own good information. It is worth noticing that two key features of PD-play are made intelligible in this example: (i) the exchange is blind (for security reasons let us suppose); and (ii) by arranging for appropriately spaced drops (once a month or once a quarter), we can assume that the problem of verificational lag can be averted.

Although our example (hardly a standard-case example) is one of iterated exchanges of information, it should be mentioned in passing that, if the exchange were strictly one-shot, with no assumption of future player interaction, then the classical PD-maneuver would indeed secure a purchase. Each official can be thought to have "everything to lose" by exchanging a true departmental secret for what turns out to be a phoney departmental secret. The minimax strategy of lying, in hopes that the other official will tell the truth (but also as a guarantee against the double-crossing worst outcome), is intelligible for our one-shot case.

But, getting back to our iterated example in which the discount parameter is high, should we expect TIT FOR TAT to be a fairly good strategy? It depends, of course, on what it plays against. Let us imagine that A plays TIT FOR TAT and B plays TESTER. Then A opens with t and B with l; at move two, A repays the defection with one of its own. TESTER apologizes and, from then on, A and B both play TIT FOR TAT, that is, they exchange good information thereafter.

Suppose, on the other hand, that B plays ALL D against A's TIT FOR TAT. Since A defects on the subsequent move in retaliation for B's defection at the previous node, then after the first move $\{A(t),B(l)\}$, A and B exchange nothing but lies. This is ironic, since the distinctive feature of ALL D (it defects at each node, no matter what) calls into play one of TIT FOR TAT's most important strategic virtues, that is, retaliation for each previous defection. The irony is that TIT FOR TAT is reduced to the same strategic untintelligibility as ALL D itself given the purposes at hand, that is, given that this is a game of "if you'll scratch my back, I'll scratch yours." The rational course is not for A to answer ALL D with a now steely and recriminating TIT FOR TAT, but to quit dealing with B and to find for himself another mole.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

We can now see how misleading it is to represent the cooperative dynamics of science in PD-terms.[7] It is quite true, as Blais convincingly explains, that cheating is efficiently discouraged in science, and that science owes its survival, certainly its high standing in human affairs, to its scrupulosity.[8] But it is doubtful that scientific scrupulosity can be made much sense of in PD-terms. The conservatism of scientific acceptance (tough peer review, high rejection rates by the good journals, etc.) is important, but it is over described in the minimax strategy of the classical game in which the best course for a author in a one-shot submission to a scientific journal is to send off a cheating piece of work, and the best course for the journal is to turn it down sight unseen. Admittedly, Blais is not tying his claim to the intelligibility of the classical strategy in real life one-shot situations of scientific disclosure, but it is instructive nevertheless to see how badly the PD-interpretation fares there. As for the iterated setup, if we take a journal's rejection as defection and its acceptance as cooperation, we have it that the top journals play something rather close to ALL D, what with rejection rates in the high ninetieth percentile. Faced with this, an author would be mad to counter with TIT FOR TAT, answering each rejection with still more of his worst junk.

Here is a final example from retail commerce. When I shop in a modern store, exchanges are frequently blind. I take it that what is in the package is what I have selected for purchase, and the store takes it that my check will not bounce. Verificational lag need not be a problem, if the store cashes the check promptly and if I inspect the contents of the wrapping without undue delay. But it is madness to represent the store as dealing with the inveterate check bouncer, an ALL D'er, by giving him TIT FOR TATly something that, time after time, he did not want, tucked away in packages that he had selected —a brick or a dead mouse, for example. A less lunatic variation of the stores TIT FOR TAT contra the bad check artist's ALL D invokes the device of "parcel pickup": the shopper "pays" the clerk with a bad check and then presents himself at parcel pickup to collect his purchase only to find that there is nothing there for him to pick up. In fairness, calling the police or refusing the defecting shopper admittance are not options available to the store, given the constraints of PD. If the store had to play its part in PD-terms, these strategies would take on some semblance of rational response. But this is just to show how inappropriate it is to impose PD–constraints in the first place.

[7] Blais, pp. 371 ff.
[8] With apologies to Professor Paul Feyerabend.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

III. CONCLUSION

I have attempted to establish that principles of cooperation in science and in life, granted that they illuminate to some extent the epistemic trust that we place, and are right to place, in the say-so of others, are not overall well-represented in PD–terms and by recourse to its winning strategies. As I have already noted, Blais is far from saying that the strategic character of cooperative epistemic regularities are either exactly or completely representable as TIT FOR TATish. He does think, however, that TIT FOR TAT is loosely representative of these cooperative norms.

I have been arguing that informational and communication exchanges, in which we are free to repose our epistemic trust, are not in general PD-like, and so that the success of TIT FOR TAT in PD-environments leaves moot its success as an explanatory device in non-PD environments. I have tried to give examples in which TIT FOR TAT is clearly an inappropriate cooperative strategy, if not worse.

Perhaps it is open to Blais to say that I have misunderstood his affection for TIT FOR TAT. After all, why is it necessary to resort to TIT FOR TAT in precisely its role as a PD-decision rule? Why not understand Blais's as a commitment to what we might call pseudo TIT FOR TAT, a strategy stripped of all its PD-trimmings and defined for any cooperative situation in which cooperation and defection can realistically be specified, and in which only the following features of (real) TIT FOR TAT are mimicked: pseudo TIT FOR TAT cooperates first, defects only to repay a defection, repays a given defection always but only once, and cooperates otherwise? So interpreted, the Blais thesis would now be that pseudo TIT FOR TAT represents and illuminates norms of epistemic cooperation.

It is well to notice that the point of Axelrod's studies is not that cooperative regularities are discernible in the communal epistemic enterprise, or that thought experiments galore can convince us of the necessity of such norms for social congress; or even that something like pseudo TIT FOR TAT is evidently sometimes at work in the dynamics of real life cooperation (in certain phases of collective bargaining, for example). Axelrod's distinctive contribution is to show, even under unrealistically severe assumptions taken to Hobbesian extremes, that, with a high enough discount parameter assumed, defective strategies cannot survive against competing cooperative strategies; that there are PD-ecologies that send defection into strategic extinction; and that this dramatic story is told in the mathematics of the PD-ecologies and is inseparable from, inheres in, the mathematics. True, the Axelrod results should be expected to

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

predict that in real life PD–ecologies we will have similar strategic outcomes, and, in particular, that something like TIT FOR TAT will do well there. What cannot be expected, just so, is that applications of Axelrod's theory will be definable for non-PD-environments (for one thing, the mathematics loses its purchase there). This cannot be expected, that is, in the absence of an appropriate representation theorem giving representation functions that take PD–ecologies into non-PD-ecologies and yet preserve the mathematics.

I stop short of saying that the representation theorem cannot be produced. I doubt that it can, but I am ready to wait and eager to see.

JOHN WOODS
University of Lethbridge and University of Amsterdam

BOOK REVIEWS

Patterns of Moral Complexity. CHARLES LARMORE. New York: Cambridge University Press, 1987. xvi + 193 p. Cloth $32.50, paper $9.95.

The title of this book does not clearly enough convey that it is mostly about the relation between moral and political philosophy, and in particular about the relation between personal moral ideals and the values and principles that ought to be realized and applied in political life. Its main contribution—and it is a significant one—lies in its defense of the liberal line on these matters: Charles Larmore argues that the state has a *specific* moral task to perform, that politics is not the realm of the *general* pursuit of moral value, and that, accordingly, law, government, and justice should be conceived and enacted in a way that is neutral between competing conceptions of what makes a life worth living.

That is the core of the book, and I shall say more about the way Larmore develops these themes in a moment. The link with the title, *Patterns of Moral Complexity,* lies in the author's more general suggestion that morality consists, not in any single value or decision procedure, but in a plurality of principles, ideals, rules, traditions, and bases for judgment, each with a distinctive contribution to make to a complicated whole.

This theme of moral complexity has several facets. Like almost everyone else who writes about moral philosophy these days, Larmore makes the rather tiresome observation that almost everyone else who writes about moral philosophy these days thinks morality is a matter of the simple application of a single master rule (like the

0022-362X/89/8606/331–333 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

utilitarian principle or Kantian universalizability) and that they neglect the role of virtue, insight, and nonalgorithmic judgment. But he does take some refreshing time out from this rather familiar plea for a more Aristotelian approach to ethics to devote a chapter to an attack on Alasdair MacIntyre's recent work, pointing out, quite correctly, how meager MacIntyre's scholarship is and how rare sound and valid arguments are in *After Virtue*.

Larmore's main difference with Aristotelianism in ethics—ancient and modern—has to do with its insistence on a *single* ordering of conceptions of human fulfilment. The premise of his liberal commitment is moral *pluralism*, i.e., the idea that there are many viable ways of living life which cannot be ordered into a single conception of the good for man. But though this sort of pluralism is common in modern ethics and political philosophy, it needs to be explicated carefully and distinguished from at least two other phenomena: irresolvable moral disagreement, and tragic moral conflict. That these latter two are not identical is shown by the fact that *A* may disagree irresolvably with *B* about the importance of moral consideration *X*, while conceding that if it *were* important it might sometimes clash tragically with some other consideration *Y* that both *A* and *B* accept. And that neither of these positions entails or is entailed by moral pluralism is indicated by the fact that *A* and *B* may both agree that they are arguing about the substance of one particular outlook on life, leaving entirely open the question of whether other conceptions of the good are available (whatever that means). Now, because of the way this book is organized, Larmore does not leave himself enough space to draw these distinctions as carefully as he might, and for a better worked-out account, readers will have to look at something like chapters 12–15 of Joseph Raz's recent book, *The Morality of Freedom*.[1]

The theme to which Larmore devotes his most sustained attention is that of liberal neutrality. His argument makes an important contribution to that debate by distinguishing two different ways in which neutrality might be conceived. It may be conceived, first, as a *modus vivendi* in politics—a way of allowing the followers of different ideals to get along together in society without having to qualify the content or wholeheartedness of their ethical convictions. Or, secondly, it may be conceived as a thesis expressing the positive ethical importance of some value or principle such as personal autonomy to which every moral person owes primary allegiance. Larmore argues, convincingly, that this "expressivist" account of the connection be-

[1] New York: Oxford, 1986.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tween neutrality and autonomy shares with its communitarian rival the conviction that ultimately our politics should express our shared commitment to the highest and most important values there are. His discussion of this theme is learned and persuasive, locating figures like Kant, Herder, Hegel, John Rawls, and Michael Sandel in a rather more sophisticated framework than we are used to in this debate.

Larmore himself opts for the *modus vivendi* approach. But his own defense of neutrality is neither cogent nor entirely consistent. To answer, e.g., the Hobbesian claim that a sovereign with authority to impose a conception of the good life is as much a *modus vivendi* as the neutrality of liberal governance, he argues that individuals with rival conceptions of the good have an interest in pursuing the matter in conversation and argument, and that neutrality can be understood largely in terms of the conditions of an "ideal speech situation" along lines sketched out by Jürgen Habermas. But this line of argument works only given certain assumptions about what moral pluralism is, and what it is for alternative conceptions of the good to confront one another in argument. As I have already pointed out, Larmore does not provide a sufficiently detailed account of pluralism, conflict, and disagreement to convince us that the proponents of rival conceptions have anything to say to one another, or that there is profit in their trying to determine through conversation who is "really" right. He certainly does not say enough to convince us that this is the *raison d'être* of liberal politics.

Once again, the contrast is with *The Morality of Freedom*. Joseph Raz has done hard work on the notion of moral pluralism, and on how it is related to themes of agreement and disagreement, on the one hand, and to autonomy and coercion, on the other. Larmore's book is provocative and worth reading, but it does not rise to the challenge of complex argument which has now been established in this area.

JEREMY WALDRON

University of California/Berkeley

NOTES AND NEWS

The Philosophical Foundation for International Thought, a nonpolitical charitable organization recently set up in Great Britain with the aim of promoting philosophical dialogue on international affairs and humanitarian issues, now offers a low-subscription membership scheme. Further details may be obtained by writing Sanjay Kewalram, School of Economic and Social Studies, University of East Anglia, Norwich NR4 7TJ, England.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

A Congress of Philosophy will take place in Nairobi, Kenya, in August 1991. Papers are invited for a symposium on the topic of Eco-philosophy. Proposed titles, along with synopses, may be sent to Hwa Yol Jung, Moravian College, Bethlehem, PA 18018.

The University of St. Andrews Centre for Philosophy and Public Affairs invites applications for the Tennent Caledonian and Royal Bank of Scotland Research Fellowships. Further information may be obtained by writing John Haldane, Centre for Philosophy and Public Affairs, University of St. Andrews, Fife KY16, 9AJ, Scotland.

The Ninth International Congress of Logic, Methodology, and Philosophy of Science will be held in Uppsala, Sweden, on August 7–14. Further information may be obtained by writing Brian Skyrms, Phil. Dept., University of California, Irvine, CA 92717.

The Center for Critical Thinking and Moral Critique is hosting the Ninth Annual, Seventh International Conference on Critical Thinking and Educational Reform on August 6–9 at Sonoma State University. In addition, intensive preconference workshops will be held on August 4–5. Further information may be obtained by writing Jane Kelsberg, Sonoma State Univ., Center for Critical Thinking and Moral Critique, Rohnert Park, CA 949248.

Loyola University of Chicago and Loyola University Press would like to announce a series of books under the general title *Values and Ethics*. The series is oriented toward values and ethics as they effect and cut across a wide range of disciplines. Abstracts and curricula vitae may be directed to Gerald Gutek, Loyola Univ. of Chicago, School of Education, Lewis Towers 850, Chicago, IL 60611.

The Philosophy Department of Hamilton College is pleased to announce the appointment of Virginia Held as Truax Visiting Professor of Philosophy for the Fall term. The Truax Visiting Professorship was endowed in 1956 by R. Hawley Truax in memory of his father, Chauncey S. Truax. Recent holders of the professorship include Joel Feinberg, Gerald Massey, Robert Holmes, Max Black, Elizabeth Flower, Abraham Edel, A. R. C. Duncan, and Richard Taylor.

The Department of Philosophy of San Jose State University is pleased to announce the appointments of Sterling V. Harwood (Cornell) and Richard Tieszen (Columbia) as Assistant Professors. Dr. Harwood will also be director of the University's Insitute for Social Responsibility.

NEW BOOKS: ANTHOLOGIES

BAIRD, ROBERT M., and STUART E. ROSENBAUM, eds.: *Morality and the Law.* Buffalo: Prometheus, 1989. 148 p. Paper $9.95.

BARTLEY, W. W., III, ed.: *The Collected Works of F. A. Hayek: vol. I: The Fatal*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Conceit: The Errors of Socialism. Chicago: University Press, 1988. xiii, 180 p. Cloth $24.95.

BATTIN, MARGARET P., et al., eds.: *Puzzles about Art: An Aesthetics Casebook.* New York: St. Martin's, 1989. xiv, 226 p. Paper $11.00.

BORRADORI, GIOVANNA, ed.: *Recoding Mathematics: The New Italian Philosophy.* Evanston: Northwestern UP, 1988. 227 p. Cloth $39.95, paper $15.95.

CERNIC, DAVID, and LINDA LONGMIRE, eds.: *The Search for Community.* Lanham, MD: UP of America, 1988. xx, 152 p. Paper $12.50.

DZIEMIDOK, B., and P. MCCORMICK, eds.: *On the Aesthetics of Roman Ingarden: Interpretations and Assessments.* Boston: Kluwer, 1989. x, 301 p. Cloth $69.00

ELLEN, ROY, et al., eds.: *Malinowski between Two Worlds: The Polish Roots of an Anthropological Tradition.* New York: Cambridge, 1988. xxv, 261 p. Cloth $39.50.

FABIAN, REINHARD, et al., eds.: *International Bibliography of Austrian Philosophy.* Amsterdam: Rodolpi, 1988. 144 p.

FAHEY, JOSEPH, and RICHARD ARMSTRONG, eds.: *A Peace Reader: Essential Readings on War, Justice, Non-Violence and World Order.* New York: Paulist, 1987. viii, 477 p. Paper $14.95.

FERRÉ, FREDERICK, ed.: *Concepts of Nature and God: Resources for College and University Teaching.* Athens: Univ. Georgia Phil. Dept., 1989. x, 258 p. Paper $7.75, spiral $8.50.

FORSYTH, MURRAY, and MAURICE KEENS-SOPER, eds.: *A Guide to the Political Classics: Plato to Rousseau.* New York: Oxford, 1989. 205 p. Cloth $39.95, paper $12.95.

FULLER, TIMOTHY, ed.: *The Voice of Liberal Learning: Michael Oakeshott on Education.* New Haven: Yale, 1989. 169 p. Cloth $20.00.

GEACH, P. T., ed.: *Wittgenstein's Lectures on Philosophical Psychology: 1946–47.* Chicago: University Press, 1988. xv, 348 p. Cloth $39.95.

GILMAN, SANDER, et al., eds. and trans.: *Friedrich Nietzsche on Rhetoric and Language.* New York: Oxford, 1989. xxvii, 273 p. Cloth $35.00.

GOTTLIEB, ROGER S., ed.: *An Anthology of Western Marxism from Lukacs and Gramsci to Socialist-Feminism.* New York: Oxford, 1989. viii, 380 p. Cloth $34.50, paper $14.95.

HARE, PETER H., ed.: *Doing Philosophy Historically.* Buffalo: Prometheus, 1988. v, 352 p. Cloth $34.95.

HOLLINGER, DAVID A., and CHARLES CAPPER, eds.: *The American Intellectual Tradition: vol. I: 1620–1865;* and *vol. II: 1865 to the Present.* New York: Oxford, 1989. xiii, 393 p; and xii, 276 p. Paper $14.95; and paper $12.95.

HOLZBERGER, WILLIAM G., and HERMAN J. SAATKAMP, JR., eds.: *George Santayana: The Sense of Beauty: Being the Outlines of Aesthetic Theory: A Critical Edition.* Cambridge: MIT, 1988. xxviii, 248 p. Cloth $35.00.

HUTCHINSON, ALLAN C., ed.: *Critical Legal Studies.* Totowa, NJ: Rowman & Littlefield, 1989. viii, 348 p. Cloth $49.95, paper $22.50.

KAELIN, E. F., and SCHRAG, C. O., eds.: *Analecta Husserliana: The Yearbook of Phenomenological Research: vol. XXVI.* Boston: Kluwer, 1989. xxxi, 445 p. Cloth $99.00.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

KRETZMANN, NORMAN, and ELEONORE STUMP, eds.: *The Cambridge Translations of Medieval Philosophical Texts: vol. I: Logic and the Philosophy of Language.* New York: Cambridge, 1988. viii, 531 p. Cloth $59.50, paper $19.95.

LIMBRICK, ELAINE, and DOUGLAS F. S. THOMPSON, eds.: *That Nothing is Known: Francisco Sanches.* New York: Cambridge, 1988. x, 310 p. Cloth $49.50.

LUFT, ERIC VON DER, ed.: *Schopenhauer: New Essays in Honor of His 200th Birthday.* Lewiston, NY: Mellen, 1988. xxxiv, 459 p. Paper $39.95.

NITECKI, MATTHEW H., ed.: *Evolutionary Progress.* Chicago: University Press, 1989. viii, 354 p. Cloth $42.50, paper $16.95.

PAPANICOLAOU, ANDREW C., and PETE A. Y. GUNTER, eds.: *Bergson and Modern Thought: Towards a Unified Science.* Chur, Switzerland: Harwood, 1988. xxi, 394 p.

PAUL, ELLEN FRANKEL, et al., eds.: *The New Social Contract: Essays on Gauthier.* New York: Blackwell, 1988. xi, 221 p. Cloth $19.95.

PAVKOVIC, ALEKSANDAR, ed.: *Contemporary Yugoslav Philosophy: The Analytic Approach.* Boston: Kluwer, 1988. vi, 319 p. Cloth $82.00.

ROGERS, G. A. J., and RYAN, ALAN, eds.: *Perspectives on Thomas Hobbes.* New York: Oxford, 1989. 209 p. Cloth $39.95.

SALLIS, JOHN C., et al., eds.: *The Collegium Phaenomenologicum: The First Ten Years.* Boston: Kluwer, 1988. 339 p.

SAYRE-MCCORD, GEOFFREY, ed.: *Essays on Moral Realism.* Ithaca: Cornell, 1988. xiii, 317 p. Cloth $42.50, paper $12.95.

SCHÜRMANN, REINER, ed.: *The Public Realm: Essays on Discursive Types in Political Philosophy.* Albany: SUNY, 1989. 309 p. Cloth $44.50, Paper $14.95.

SINNOTT, JAN D., ed.: *Everyday Problem Solving: Theory and Applications.* Westport, CT: Praeger, 1989. xvii, 336 p. Cloth $49.95.

SLATER, JOHN G., ed.: *Bertrand Russell: Essays on Language, Mind and Matter: 1919–26: The Collected Papers of Bertrand Russell.* London: Unwin Hyman, 1988. xl, 658 p.

SOLOMON, ROBERT C., and KATHLEEN M. HIGGINS, eds.: *Reading Nietzsche.* New York: Oxford, 1988. xiv, 251 p. Cloth $29.95.

WINCHESTER, IAN, and KENNETH BLACKWELL, eds.: *Antinomies & Paradoxes: Studies in Russell's Early Philosophy.* Hamilton, Ontario: McMaster UP, 1989. 248 p.

WINNEFRITH, TOM, and CYRIL BARRETT, eds.: *The Philosophy of Leisure.* New York: St. Martin's, 1988. ix, 174 p. Cloth $39.95.

ZILBERMAN, DAVID B., and ROBERT S. COHEN, eds.: *The Birth of Meaning in Hindu Thought.* Boston: Reidel, 1988. xxii, 368 p. Cloth $99.00.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVI, NO. 7, JULY 1989

LEIBNIZ, KEYNES, AND THE RABBIS ON A PROBLEM OF DISTRIBUTIVE JUSTICE

Leibniz' fascination with difficult legal issues dates from his early years and found expression in the doctoral dissertation *De casibus perplexis in jure,* which he presented to the University of Altdorf at the age of twenty in 1666. And this interest in law intersected with his concern for combinatorics and probability when, in a 1687 letter to Vincentius Placcius,[1] one of his longest-term correspondents, he wrote as follows:

> If two litigants lay claim to a sum of money, and if the other claim of the one is twice as probable as that of the other, the sum should be divided between them in that proportion (*ibid.*, p. 36).

In seeking to follow this prescription, there is, of course, the obvious difficulty of assessing, relative to the evidence at hand, the actual mathematical probability that a particular party is the valid claimant. Leibniz himself recognized, and indeed stressed, that what matters for probability is not the mere number of witnesses speaking for one claimant or the other, nor again the mere number of arguments favoring one side against the other, but rather the overall weight of such evidential considerations.[2] And he regarded the resultant task

[1] Leibniz' Latin letter is printed in his *Opera omnia,* Ludovico Dutens, ed. (Geneva: Fratres de Tournes, 1768); see vol. VI, pt. 1. It is quoted by J. M. Keynes in *A Treatise on Probability* (London: MacMillan, 1921), p. 311. Leibniz refers to Placcius as "the famous jurisconsult (*Ictus*) of Hamburg, whose learning, industry, profoundity and unusually good intentions I esteem highly," in a letter to Gabriel Wagner of late 1696, in *Philosophische Schriften,* C. I. Gerhardt, ed., vol. VII (Berlin: Wiedmann, 1890), p. 518.

[2] See Louis Couturat, *La Logique de Leibniz* (Paris: Alcan, 1901), p. 247, no. 5. See also *New Essays,* bk. IV, ch. ii, sect. 14; Academy edition, P. Remnant and J. Bennett, trans. (New York: Cambridge, 1981), p. 372. (All my quotations from the *New Essays* [hereafter NE] are from this translation.)

0022-362X/89/8607/337–352

© 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

of devising an instrumentality for assessing this weight of evidence as a prime desideratum: "I maintain that *the study of the degrees of probability* would be very valuable and is still lacking, and that this is a serious shortcoming in our treatises on logic" (NE, bk. IV, ch. ii, sect. 14, p. 372).[3] And again:

> *Endoxon,* for Aristotle, is whatever is accepted by the greatest number or by the most authoritative. . . . But probability or likelihood is broader: it must be drawn from the nature of things; and the opinion of weighty authorities is one of the things which can contribute to the likelihood of an opinion, but it does not produce the entire likelihood by itself. Even at the time when Copernicus was almost alone in his opinion, it was still incomparably more *likely* than that of all the rest of the human race. I suspect that establishment of an *art of estimating likelihoods* would be more useful than a good proportion of our demonstrative sciences, and I have more than once contemplated it (NE, bk. IV, ch. ii, sect. 14, p. 372/3).[4]

With respect to this important desideratum, we are not even yet home free. The valiant efforts of J. M. Keynes and Rudolf Carnap and their successors in the present century notwithstanding, the problem of devising a quantitative measure of probability in the specific sense of weight of evidence remains to this very day substantially unresolved. But, of course, this difficulty about measuring evidential probabilities only relates to the matter of the *practicability* of implementing the principle of proportionate division. The question of its *appropriateness* remains untouched. Just what are its merits from the standpoint of the right and the just? Is this rule of evidence-proportionate division indeed one that we should want to use when and where we can?

It is this question of desirability (rather than that of feasibility) which will primarily concern us here. For it will be argued in the subsequent discussion that the plausible appeal of Leibniz' endorse-

[3] Cf. also pp. 206, 465/6. And see also the letter to Thomas Burnet of February 1697, in *Philosophische Schriften*, vol. III, C. I. Gerhardt, ed. (Berlin: Wiedmann, 1887), pp. 186–197; see pp. 193/4. For the legal background of Leibniz' interest in probability, see Ian Hacking, *The Emergence of Probability* (New York: Cambridge, 1975), ch. 10. The historical stagesetting of early probabilism is informatively surveyed in Lorraine Daston, *Classical Probability in the Enlightenment* (Princeton: University Press, 1988).

[4] The most comprehensive discussion of Leibniz on probability still remains that of Couturat, *op. cit.*, ch. VI. Daston has maintained that "[t]his shift of emphasis [in Jakob Bernoulli's 1713 *Ars conjectandi*] from expectations to probabilities . . . brought mathematical probability theory a new set of applications to belief and credibility, inaugurated by Bernoulli himself" (*op. cit.*, p. 34). But these applications had already figured in the mind of Bernoulli's friend and correspondent, Leibniz.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ment of evidence-proportionate division is something highly problematic which rests on assimilating the judicial allocation problem to a rather similar looking, but actually very different issue—the problem of fair division in gambling. Let us consider how this is so.

I.

Suppose that a game of dice has reached a stage where only two players, A and B, remain as contenders, and where winning is to be decided one way or the other by A's getting an odd number greater than 2 on a single remaining die toss. The outcome then hinges on A's getting 3 or 5. But suppose further that the game has to be broken off at this very point. How should the stake constituting the "pot" now be divided between A and B?

Since A's winning has the probability 1/3 and B's the probability 2/3—so that A's chances of winning are exactly half of B's—it seems proper and sensible to divide the stake between them in the proportion 1:2. The guiding idea is that a just division can be based on exactly the proportionality principle that was contemplated in Leibniz' contention:

$$\frac{\text{proportionate share of } A}{\text{proportionate share of } B} = \frac{A\text{'s probability of winning}}{B\text{'s probability of winning}}$$

On this basis, the following relationship holds with respect to any one of the participants X:

$$\frac{\text{proportionate share of } X}{X\text{'s probability of winning}} = \text{constant}$$

In consequence,

$$\text{proportionate share of } X \sim X\text{'s probability of winning}$$

Since the several proportionate shares must, in the aggregate, sum to the total stake at issue, we obtain:

$$\text{proportionate share of } X = X\text{'s mathematical expectation}$$

Thus, everyone is treated just the same: everyone's share is just exactly their probabilistic expectation. In this sense, the division is perfectly fair; a principle of equity is in operation.[5]

On this basis, then, a participant's appropriate share in the overall stake is simply the probabilistically determined mathematical expectation (expected value) of the gamble for this particular party. And

[5] Though a professional lawyer, Leibniz, like Pascal, sought to get at fair contracts via probability, while Huygens and DeWitt wanted to do it the other way around. See Daston, *op. cit.*, pp. 28/9 and passim.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

this in fact is the traditional resolution of the gambling division problem, of the "problem of points" of sharing out the stake between the participants on a fair game that is stopped in mid-career, a resolution which Leibniz enthusiastically endorsed.[6]

It was this general issue which provided the launching impetus to the development of the calculus of probability. Gábor J. Székely[7] recently summarized the historical situation as follows:

> This paradox was first published in Venice in 1494 in a summary of the mathematics of the Middle Ages. The author, Fra Luca Paccioli (1445–1509), entitled his book *Summa de arithmetica, geometria, proportioni et proportionalità.* This book first uses the word *million* and explains the rules of double entry for the first time. It is interesting to note that Fra Luca and Leonardo da Vinci became close friends in Milan and, due to this friendship, Leonardo illustrated Fra Luca's work *De Divina Proportione,* which was published in Venice in 1509. Oystein Ore recently found an Italian manuscript dating from 1380 which also mentions the paradox of division. Many a thing indicates that the problem is of Arabic origin, or at least reached Italy through Arabic teaching. . . . An incorrect solution was given by Niccolo Tartaglia (1449–1557), though he was such a genius that he discovered the formula to solve cubic equations in one night in a mathematical duel. After several unsuccessful attempts, Pascal and Fermat eventually gave the right answer to the problem independently of each other in 1654. It was such an important discovery that this date is considered by many people to be the birth of probability theory, and all the previous results belong only to its prehistory (*ibid.*, pp. 9–10).

Here is how Leibniz himself described the general situation in 1705:

> Mathematicians have begun, in our own day, to calculate the chances in games. It was the Chevalier de Méré—a man of acute mind, a gambler and philosopher, whose *agreements* and other works have been published—who prompted them by raising questions about the division of

[6] Note that Leibniz does not, however, think that expected values are inevitably a decisive guide to assessing the acceptability of gambles. "Un homme qui a 20 000 écus de bien, ne doit pas le hazarder en un seul coup contre 100 000 écus, car ces 100 000 gagnés n'augmenteront pas beaucoup son bonheur, et les 20 000 perdus le rendront miserable" (*Opuscules,* Couturat, ed., pp. 182/3). This line of thought prefigures the distinction between mathematical and moral expectation which underpins David Bernoulli's 1738 proposal for resolving the St. Petersburg Paradox. And it anticipates D'Alembert's question that, given the choice between a .99 probability of 1,000 écus and a .01 probability of 99,000 écus, "where is the man so foolish as to prefer the offer of 99,000 écus? The *expectation* in the two cases is not *really* the same; even though it is the same according to the rules of probability" (*Opuscules mathématiques* (Paris: David, 1761–80), vol. 4, p. 83). Cf. also the discussion of this issue in my *Risk: A Philosophical Introduction to the Theory of Risk Evaluation and Management* (Washington: UP of America, 1983), pp. 70–76.

[7] *Paradoxes in Probability Theory and Mathematical Statistics* (Boston: Reidel, 1986).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

stakes, wanting to know how much [a given player's part in] a game would be worth if this game were interrupted at such a point. Accordingly, he invited his friend M. Pascal to take a brief look at the problem. The question caused a stir and prompted M. Huygens to write his treatise on chance. Other learned men joined in. Certain principles were established, and were also employed by Pensionary De Witt in a little discourse, published in Dutch, on annuities. The foundation they built on involved *prosthaphaeresis,* i.e., arriving at an arithmetic mean between several equally admissible hypotheses. Our peasants have used this method for a long time, guided by their natural mathematics. For instance, when some inheritance or piece of land is to be sold, they appoint three teams of assessors—these teams are called *Schurzen* in Low Saxon—and each team assesses the commodity in question. Now suppose that the first estimates its value at 1000 crowns, the second at 1400 and the third at 1500; they take the total of these three, which is 3900, and because there were three teams they take a third of this, 1300, as the mean value sought. Or, what comes to the same thing, they take the sum of one third of each estimate. This is the axiom: *aequalibus aequalia*—like hypotheses must receive like consideration. But when the hypotheses are unlike, we compare them with one another. Suppose, for instance, that with two dice one player will win if he throws a 7 and the other if he throws a 9. We want to know their relative likelihoods of winning. I say that the second player is only two thirds as likely to win as the first player, since there are three ways in which the first can throw a 7 with two dice—1 and 6, or 2 and 5, or 3 and 4—whereas there are only two ways in which the second can throw a 9, namely by throwing 3 and 6, or 4 and 5. And all these ways are equally possible, so that the likelihoods, which match the numbers of equal possibilities, will be as 3 to 2, of 1 to 2/3. I have said more than once that we need a new kind of logic, concerned with degrees of probability, since Aristotle in his *Topics* could not have been further from it: he was content to set out certain familiar rules, arranged according to the commonplaces—rules which may be useful in some contexts where a discourse has to be developed and given some likelihood—without taking the trouble to provide us with balances which are needed to weight likelihoods and to arrive at sound judgments regarding them (NE, bk. IV, ch. XVI, sect. 9, pp. 465/6).

In his discussion of "mathematical expectation," which, he maintains, "is a technical expression originally derived from the scientific study of gambling and games of chance, and stands for the product of the possible gain with the probability of attaining it," Keynes went so far as to credit Leibniz with originating this conception (*op. cit.*, p. 311). But this attribution is not really accurate, since the idea had certainly been used earlier by mathematical theorists in France (de Méré, Pascal, Fermat) and Holland (Huygens, de Witt)—in connection both with games of chance and with the economics of insurance. It would be more correct to say that, although Leibniz did not initi-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ate this idea, he extended and generalized it beyond the level of earlier treatments, insisting that one should recognize this idea of expectation as an instrument of vastly diversified applicability and utility.

As regards Leibniz' formulation of the issue, a prime text is his essay *De incerti aestimatione,* where he puts the salient point as follows:[8]

> If among all possible outcomes some yield the value A, others the value B, and the rest the value C, then the total expectation (*spes*) will be the sum of the several values multiplied by the number of possible outcomes that yield them, divided by the number of all possible outcomes. Thus if the number of outcomes that can yield the value A is a, the number of outcomes that can yield the result B is b, and the number of outcomes that can yield the result C is c, and the number of all possible outcomes is n, then the expectation will be
>
> $$s = \frac{aA + bB + cC}{n}$$

Here supposing equivalent outcomes—so that a/n is the probability of realizing the value A, b/n that of realizing B, and c/n that of realizing C—Leibniz' treatment of the issue makes it clear that expectation is the aggregate sum of the product of the probabilities of the several outcomes by their respective values.

Thus—to take an insurance example—if the probability of a ship's sinking in a storm during a certain voyage is one in a hundred, with the loss standing at 10,000 ducats on the outbound voyage and 40,000 ducats on the return voyage, then the overall effective magnitude of this misfortune of "loosing the ship in a storm" would be appraised at

$$(.01 \times 1/2) \times 10{,}000 + (.01 \times 1/2) \times 40{,}000$$

or 250 ducats, which thus represents a reasonable measure of the magnitude of the negativity at issue ("a possible loss of our ship in a

[8] *Opuscules et fragments inédits de Leibniz,* Couturat, ed. (Paris: Alcan, 1903), pp. 569–570. "Si ex omnibus eventibus aliquot dent rem A, aliquot alii rem B, et reliqui rem C, erit spes tota aggregatum ex rebus singulis in numerum eventuum qui eas dare possunt ductis, divisum per numerum eventuum possibilium omnium. Ut si numerus eventuum qui dare possunt rem A sit a, numerus eventuum qui dare possunt rem B sit b, et numerus eventuum qui dare possunt rem C sit c, et numerus omnium eventuum sit n, erit spes s aequ $\frac{aA + bB + cC}{n}$." The connection between hope (*spes*) and probability goes back to Locke: "*Hope* is that pleasure in the mind, which everyone finds in himself, upon the thought of a probable future enjoyment of a thing which is apt to delight him" (*Essay,* bk. II, ch. 20).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

storm"), and affords a quantity to guide us in assessing the cost of insurance.

It is interesting to see Leibniz' fertile mind at work envisioning various applications for his favored evaluation principle which are remote from its origin in the division of gambling stakes and the pricing of insurance. Thus, in the long passage cited above, he applies it to quantitative appraisal. Assuming that each appraiser is equally competent, and so has equal chances of giving the correct appraisal, we should simply take the arithmetical mean of their estimates. And Leibniz also contemplates using the same approach with respect to medical prescriptions. If various (*ex hypothesi* equally competent) physicians prescribe varied doses of a medication in a given case, we should again adopt the arithmetical mean.[9]

As such examples show, Leibniz was prepared to move well beyond the close connection between expectation and equity which otherwise characterized the thought of the other founders of probability theory. In fact, he sought to extend this approach in many directions. For example, he proposed to apply it in moral casuistry, subject to the idea that the appropriateness of an action will depend on two things: (1) "safety"—its inherent moral appropriateness/inappropriateness; and (2) "probability"—the likelihood of its being mandatory; the degree to which its requiredness can be established. Thus, if an action has a low degree of safety because it is inherently morally questionable (e.g., abandoning one's family to lead a life of meditation and prayer), then a stronger rationale of validation will be needed to validate its appropriateness:

> And when our moralists—I mean the wisest of them, such as the present General of the Jesuits—bring in the question of what is safest as well as of what is most probable, and even put safety ahead of probability, they do not really abandon the most probable. For here the question of *safety* is the question of the *improbability of an impending evil.* Moralists who are lax about this have gone wrong largely because they have had an inadequate and over-narrow notion of probability, which they have confused with Aristotle's *endoxon* or *acceptability:* in his *Topics* Aristotle aimed only to conform to the opinions of other people, as did the orators and the Sophists. *Endoxon,* for him, is whatever is accepted by the greatest number or by the most authoritative; he was wrong to restrict his *Topics* to that, and this approach meant that he only concerned himself there with accepted maxims, most of them vague—as though wanting to reason by means of nothing but old jokes and proverbs. But probability or likelihood is broader—it must be drawn from the nature of things; and the opinion of weighty authorities is one of the things which can contribute to the likelihood of an opinion, but it does

[9] *Opuscules,* Couturat, ed., p. 226; cf. NE, bk. IV, ch. XVI, sect. 9, p. 465.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

not produce the entire likelihood by itself (NE, bk. IV, ch. ii, sect. 14, p. 372).[10]

In fact, Leibniz proposed to use this approach in measuring the overall value of goods and evils within a broadly conceived moral calculus. Effectively, the magnitude of something good or bad should be measured by the *product* of its inherent value by the probability of its realization:

(inherent value) × (probability of eventuation)

He observes:

> The question of how inevitable a result is (*la grandeur de la consequence,* i.e., its likelihood) is heterogeneous from—i.e., cannot be compared with—the question of how good or bad it is (*la grandeur du consequent,* i.e., its value). So in trying to compare them, moralists have become muddled, as can be seen from writings on probability. The fact is that in this as in other assessments which are disparate, heterogeneous, and with more than one dimension (so to speak), the magnitude of the thing in question is made up proportionately of two estimates; as with a rectangle . . . (NE, bk. II, ch. xxi, sect. 66, pp. 205/6).

In the case of a disjunctively composite development, which can be realized in a plurality of distinct ways, we are to assess its overall value as the appropriately weighted arithmetical mean of the values of its components in just the way that was explained above. That is, the mathematical expectation is to be our measure. As Leibniz saw it, the expected value idea of a multiplicative contingency of probability and magnitude affords a tool of inverse veribility and power for the assessment of goods and evils in general. Where the early prohibitants conceived of an instrument for dealing with games of chance matters of insurance, Leibniz, with his characteristic expansiveness of view, envisioned a generalized moral calculus for rational deliberators in situations of uncertainty.

II.

Commenting on the principle of proportionate division in deciding legal conflicts, Keynes joined Leibniz in endorsing it with the observation: "The doctrine seems sensible, but I am unaware that it has ever been acted on" (*op. cit.*, p. 311, fn. 1). But are Leibniz and Keynes indeed right in claiming that it is sensible to use expectation as a guideline for judicial award?

Someone might object straightaway that the principle of propor-

[10] Again: Bona malave aestimanda sunt separatim tum ex magnitudine sua, tum ex probabilitate. . . . Et, si aequalia sint erunt in ratione probabilitatis, si aeque probabilia, in ratione magnitudinis. Et si inaequalia et inaequaliter probabilia sint, erunt in ratione composita magnitudinis et probabilitatis [Leibniz, *Vorausedition,* Faszikel 6 (Muenster: Leibniz-Forschungsstelle, 1987), pp. 1136/7].

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tionate division is unworkable, because all too often the contested object cannot be physically divided. (We would not, for example, contemplate cutting up a husband for division between two rival claimant wives.) But, of course, what we could divide in the probability of its allotment in a duly proportioned lottery. And this is in fact how the matter is handled in certain real-life situations. For example, when precedence must be given to one of two ambassadors whose claims are equivalent, or when the opening play of a game must be given to one of two competing teams, the matter is standardly settled by a coin toss. As such considerations indicate, the indivisibility objection does not really militate against the principle of proportional division. But other, more telling objections in fact exist. To clarify their nature, it is useful to look at the issue from the vantage point of some distinctions.

III.

Two different approaches to distribution situations can plausibly be contemplated: the one based on proportionality, the other on predominance. The proportionist approach makes a fractional allocation in line with the comparative strength of the claim that a party advances. The predominantist approach makes a total allocation in favor of whatever claim is the strongest in the scale of relevant comparison. On this approach, the stronger claim "tips the balance" in favor of its possessor: the side whose weight is comparatively the heaviest, by however little, will prevail all out.

Traditional legal practice firmly opts for this leader-takes-all line. Virtually every legal system proceeds by way of determining the strongest claim and makes its award strictly on this basis, setting the claims of the lesser claimants at nought. In view of this situation, it is nowise strange to find Keynes conceding, as we have seen, that the proportionist principle that he joins Leibniz in favoring is such that: "I am not aware that it has evern been acted on." Established legal practice is everywhere squarely on the side of predominantism. In politics, to be sure, we find both proportional representation and leader-take-all election systems. But legal as opposed to political systems steer clear of proportionism. To be sure, this circumstance does not of itself settle the issue of whether this approach is just, fair, and reasonable. The question remains: Does the deeper justice of the matter lie with Leibniz and Keynes in their advocacy of expected-value proportionism?

A strong case can in fact be made against Leibniz and Keynes and in favor of the traditional legal approach. But to see this it is necessary to distinguish further between two importantly different sorts of distributive allocation situations, namely, that of ownership *creation* with respect to heretofore unowned assets, and that of ownership

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

determination in a situation of evidential insufficiency (uncertainty) concerning pre-existing ownership. The former is a matter of the accommodation of competing *entitlement* considerations in a no-ownership situation. The latter is a matter of the adjudication of conflicting *evidential* considerations in a situation where an ownership relation is already present. Two very different things are at issue here.

The issue of ownership creation is typified by the previously contemplated situation of dividing the stake when a dicing game must be prematurely terminated—exactly the sort of problem which in fact gave rise to the calculus of probability. In such cases of game termination, there is as yet no ownership of a particular item—there are simply variable claims, entitlement considerations of variable strengths. The division that is being made is designed to initiate possession and create ownership. The controlling factor is simply that of fairness to the competitors. A suitable legal example might be that of the decedent who died leaving his estate to his domestic servants in otherwise unspecified proportions—some beneficiaries having many years service, others having recently joined the staff. Exactly as with the gambling case, each party here has a claim on some assets that neither as yet actually owns. In this sort of case, a proportional division by strength of claim—say, by length of service—seems perfectly equitable; exactly because it treats everyone alike in allocating to each contesting party a share proportionate to the strength of its actual claims. Where there are as yet no pre-established ownership rights upon the scene, but only a plurality of competing claims to be accommodated, it seems perfectly reasonable to effect a proportionate division in line with the respective strength of these variable claims.

A very different sort of situation arises, however, where a proprietary interest already exists, but there merely is evidential uncertainty regarding the identity of the owner. Take, for example, the situation when two different John Smiths claim a lost-and-found item known to belong to a John Smith. Or take Leibniz' own example of rival claimants to ownership of a certain item of value—for example, when there is some question about to which of two parties a tract of land has actually been conveyed. In all such cases, the actual practice in legal systems of every sort is to recognize the ownership of the party whose case is the strongest. And this approach is standard in many other contexts as well—for example, with allocating credit for geographic or scientific discoveries (as in the dispute of Peary versus Cook as claimants to the discovery of the North Pole). We standardly treat all such cases as a form of existing ownership and do not divide the credit.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

And this seems reasonable and proper, because in these matters of ownership determination there is in fact excellent reason for taking the leader-takes-all approach. To see this more clearly, consider the following sort of situation:

> Let it be that there is an item of substantial value V whose possession is contested between two parties, A and B, in circumstances where no one else has any plausible claim to this item. And let it be that, relative to the evidence at hand, the case for crediting this item to A is somewhat stronger than the case for crediting it to B. Then how, by rights, should the total value V be divided between A and B?

Suppose that we divide the contested value between A and B in the proportion x:$1 - x$ for some as yet undecided value x. Now, if it actually belongs to A, then its owner has been shortchanged by the quantity $1 - x$. On the other hand, if it actually belongs to B, then its owner has been shortchanged by the quantity x. Let us further specify:

p = the probability that the contested item actually belongs to A

$1 - p$ = the probability that the contested item actually belongs to B

(Note that by assumption $p > 1/2$.) On this basis, we can now calculate the overall expectation of shortchanging the true owner of the item under contention:

$$p(1 - x) + (1 - p)x = p + x - 2px = p + x(1 - 2p)$$

Since $1 - 2p$ is *ex hypothesi* negative, this quantity is clearly minimized (to $1 - p$) by setting $x = 1$. Accordingly, in order to minimize the expectation of inappropriate deprivation, we must follow the all-out rule and allocate the whole of the credited item to the party whose claim is the strongest.

Note specifically that, if we made the allocation in line with the proportionality principle in the ratio p:$(1 - p)$, that is, effectively letting $x = p$, then we would fix the value of the expectation of shortchanging at:

$$2p - 2p^2 = 2p(1 - p)$$

Since $p > 0.5$, so that $2p > 1.0$, this quantity is certainly not less than the value $1 - p$ that we secure by setting $x = 1$. Accordingly, proportionate allocation unquestionably increases the prospect of an injustice. The predominantist leader-take-all method is demonstrably preferable in contexts of contested ownership.

And it is also helpful to consider the matter is a somewhat different perspective. Suppose in a case of the preceding sort that we were to

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

allocate the item to A or B probabilistically (by lottery or the like) with

$$q = \text{pr (it goes to } A\text{)}$$

$$1 - q = \text{pr (it goes to } B\text{)}$$

Then the overall expectation of an *injustice* is:

$$p(1 - q) + (1 - p)q = p + q - 2pq = p + q(1 - 2p)$$

Since $1 - 2p$ is negative (because *ex hypothesi* $p > 0.5$), we minimize this quantity with $q = 1$. From this perspective, too, it emerges that we must make our allocation all out in line with the stronger claim to minimize the expectation of injustice. (The proportional allocation effected by letting $q = p$ does *not* fill the bill.)

As these considerations indicate, in cases of ownership determination the sensible standard of minimizing the expectation of short-changing the proper owner in giving him less than is his just due underwrites the leader-takes-all methodology reflected in the traditional legal approach.[11]

IV.

In his tract on equity,[12] Leibniz affirms that justice (*justitia*) is a matter of *congruitas ac proportionalitas quaedam*. And there is no doubt that this idea that justice calls for implementing the maxim "according to each what is his own" (*suum cuique tribuere*) and giving everyone "his proper share" goes far back in the history of European thought on the issue.[13] The rule "to each according to his claims"—the "Canon of Claims," as it has been called—is certainly an acceptable principle of distributive justice.[14] But its implementation take very different forms. On the one hand, it can take the proportionalist line of a division according to the relative evidential strength of each claim. But, on the other hand, it can take the form of a distribution made on the more generalized basis of taking everyone's claims into appropriate account without necessarily effecting a fractional resolution on the basis of evidence-proportional shares.

It is suggestive in this regard to return once more to Aristotle's

[11] But what of the case where one does not know whether an ownership relation exists to warrant a property claim—say, whether the ring that hikers find in a field was lost or thrown away by its owner? Then, clearly, the ethical thing to do would be to presume ownership until the evidential situation indicates otherwise. The nature of the object would be relevant here, however—if golden nuggets were found by the side of a stream in a gold-yielding region, it would be implausible to assume that a prospector had lost them there.

[12] *Juris et aequi elementa*, in *Mitteilungen aus Leibnizens ungedruckten Schriften*, Georg Mollat, ed. (Leipzig: Haessel, 1893), p. 22.

[13] See Plato, *Republic*, I, 6, 331E; and Aristotle, *Nicomachaen Ethics*, V, 5.

[14] See my *Distributive Justice* (Indianapolis: Bobbs-Merrill, 1966), pp. 81–83.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

important distinction between *distributive* and *corrective* justice.[15] The former is concerned with the distributional division of heretofore unallocated assets held as common property. The latter is concerned not with a distribution at all, but with the restoration of a *status quo ante* in situations of pre-existing ownership. Now, in cases of the entitlement accommodation at issue with no-ownership situations, distributive proportionalism makes eminent sense. But, in cases where a relation of ownership is already in place as a somehow accomplished antecedent fact, it can be seen that proportionalism is inappropriate from a moral point of view. For proportionalism makes for injustice in situations of pre-existing ownership.

An important principle lies in the background here. The failure of proportionality to meet the demands of justice in situations of pre-existing ownership is clear and intuitive. If a group jointly finds a long-buried treasure, it is only fair that they should all share such a treasure-trove alike. But, if the "find" is something that actually already belongs to one of them—having been lost a matter of days or months ago—then that is where it should go. As far as the rights and wrongs of the matter goes, established ownership makes a decisive difference. Once ownership is an accomplished fact—once there has been some act or transaction that gives somebody a due claim to ownership—then justice requires us to do the very best we can to acknowledge this circumstance.

Of course, one can quarrel about what constitutes ownership. And one can even argue about its very desirability. Perhaps "bourgeois" principles of ownership are regrettable; perhaps "property is theft" and the aggregate wealth of a community should be seen as a "social product" available for proportionate redistribution. Let all that be as it may. Nevertheless, once actual ownership is there as an accomplished fact to govern people's expectations and understandings in line with the duly established social ground rules that prevail at the time, then abstract justice calls for its recognition. Pascal put the salient point in a way at once stark but true: "Nothing is just merely by abstract reason alone. . . . The arbiter of equity is custom—simply in virtue of its being generally received."[16] Accordingly, there is a harsh truth in Anatole France's caustic observation that "[j]ustice is made to give everyone his due; to the rich his riches, to the poor his poverty." That is simply the sort of thing that Aristotelian corrective justice is. One cannot fairly change the rules in mid-game: once ownership claims are in place through the entitlement-establishing

[15] *Nicomachean Ethics*, v, 2–4.

[16] "Rien suivant la seule raison n'est juste de soi. . . . La coutume est toute l'équité, par cette seule raison qu'elle est reçue" (*Pensées*, sect. 60/294; cf. 525/325).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ground rules that determine the expectations of people in the prevailing communal order, we cannot without violations of justice cancel all this out and return to an antecedent state of nature in which the playing field is leveled. The established ownership practices of a social order define entitlements and expectations in a way that cannot be ignored without the injustice of doing violence to warranted claims. Considerations of policy may at times validly override such claims (think of "eminent domain"), but they cannot simply be omitted from the social calculus.[17] Thus, once something actually belongs to somebody, on the basis of the expectation-determinative principles that prevail in the community, the wheels of justice will have to turn in a consonant direction. Antecedently, the person who wins the lottery or the game has no greater claim on the stake than any of the rival competitors. (She certainly has no particular claim on winning.) But once she has won, fair and square (though doubtless "undeservingly"), her claim on the prize is absolute in the sight of justice. The ethics of desert does not come into it at this level. In life's lottery as in the state's, the winner need not deserve the prize more than another; but once it has been fairly won, its ownership becomes absolute in the sight of justice. Justice is something *legalistic* rather than *moralistic;* Cicero was right in saying that "the origin of justice is to be found in the law."[18]

As this perspective on the matter indicates, the error committed by Leibniz—and also by Keynes in endorsing his proportionalism as "sensible"—lies in assimilating the case of ownership determination at issue in situations of Aristotelian corrective justice to that of ownership creation at issue in situations of Aristotelian distributive justice. With ownership creation, the proportionalist approach is altogether fair and proper, seeing that a proportional allocation recognizes the variable strength that competing claims exert in an ownership-indeterminate situation. Clearly, when a division is to be made in the absence of any pre-existing ownership entitlements, then proportionality claims is the method of allocation which does least violence to the principles of justice by putting each claimant into the same equitable everything-considered position. But with ownership determination the matter stands on a very different footing. Here any sort of proportional division (be it by division into shares or by a probabilistic allocation) has the defect of treating all

[17] Social theoreticians are reluctant to face the inconvenient fact that no prevailing system of distribution—however imperfect—is automatically shown to be inappropriate because of those imperfections, since there is always the prospect that the imperfections encountered in realizing each of the available alternatives are yet worse.

[18] *De legibus,* I, vi, 19.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

claimants uniformly in a situation that is not uniform, since by hypothesis one of them, unlike the others, is actually the owner of the contested item. In such situations of pre-existing ownership, the salient issue is not so much fairness as one of a rigorously corrective justice that must give ownership entitlements their due. Once ownership is duly established under the aegis of a society's prevailing rules, it is a thing to be reckoned with, a factor that cannot—in simple justice—simply be brushed aside as an irrelevant inconvenience. And here the interests of justice are, as we have seen, best served by the traditional, legalistic approach of recognizing the strongest claimant as proper owner.[19]

V.

Interestingly, the issue was viewed in exactly the right perspective by those masters of legalistic subtlety, the medieval Talmudists. For consider:

> The Talmud tells of a man who learned that of the ten sons his wife bore, he was the father of only one. Before his death, he bequeathed his entire estate to his only son, without specifying who he was. The case came before Rabbi Bana'ah who instructed the ten claimants. "Go beat on your father's grave until he will arise and reveal to you to whom he left his [property]" (*Bava Batra*, 58a). Only one of the sons refrained from such disrespectful behaviour, and so Rabbi Bana'ah awarded the bequest to him. Rabbi Asher in the course of a responsum dealing with verdicts based on probabilities comments: "On the strength of this . . . we do not say 'Let it stay till Elijah comes' [that is—let judgment be suspended and the property be held in escrow by the court]; rather let him judge in accordance with what his eyes behold, a probable opinion. On a small probability—that it appeared to him that the true son had respect for his father firmly implanted in his heart and paid him honour—he gave him all the property" [Rabbi Asher ben Yehiel, *Responsa* 107:6 (p. 196, col. 2)].
>
> Here we have a case where apparently all the claims are initially of equal probability. The test devised to ascertain who is the true son is certainly not a very decisive one: Rabbi Asher emphasizes that it is only a "small probability." Yet because one claim has a higher probability than any of the others, we are justified in reaching a decision.
>
> It ought to be pointed out that this case is quite unlike that discussed above . . . [when a husband and wife were both killed in the collapse of a house, the question arose whether his or her heirs are to inherit in the fact of a contention about who died first] where the ruling is that "property that is in doubt is divided equally." The Talmud distinguishes be-

[19] The difference rests on a theoretical basis. With pre-established ownership, the missing information that prevents our making an ideally correct award relates to the accomplished past; with gambling division it relates to the contingent future. There being no fact of ownership here, we cannot see the available information as playing an evidential role.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tween a dispute where all the conflicting claims have equal probability but only one has true validity, and a case in which all claims may in fact be equally valid, so that an equal division need not necessarily be in error. In the latter case, equal division is indicated, but in the former, since the entire property really belongs to one litigant only, if the evidence is equally balanced, no verdict can be rendered and the property must be held in escrow "till Elijah comes" (*Bara Metza,* 3a).[20]

The acute rabbis got the matter exactly right. On the one hand, in the case of the two conflicting groups of heirs contesting a legacy, there is no pre-existingly valid fact of ownership, and a proportional division is perfectly right and proper. On the other hand, in the case of the unidentified son, the problem is one of acknowledging the claims of the one and only valid heir, and an allocation in line with the single maxiprobable claim is quite in order. The Rabbinic analysis is right on target in its differential approach to cases of pre-existing claims as contrasted with cases of claim creation.

VI.

The upshot of our deliberations is clear. In assimilating the situation of a legal dispute between rival claimants to ownership of an item of value to the quite different situation of dividing the stake in games of chance, Leibniz committed an oversight. For in the latter case, use of the proportionality principle is indeed patently sensible and morally appropriate. But, in the former situation, proportionality is inappropriate and violates the demands of distributive justice. Leibniz' enthusiasm for the expected-value approach seems to have carried him too far into endorsing the use of expected values as an allocation mechanism in situations where this is not a valid approach.

But observe an interesting and noteworthy circumstance. The arguments that clearly show the inappropriateness of the expected-value standard of proportionate shares in this contested-claim case themselves turn on use of an expected-value assessment, proceeding by means of expected-value comparisons in their resort to minimizing the expectations of shortchanging people and making unjust awards. Interestingly, the inappropriateness of allocation by particular expectations is thus shown by expected-value considerations.

In the end, then, our analysis reinforces rather than impairs Leibniz' fundamental insight that expected values are crucial standards of rational appraisal in matters of distributive justice.

NICHOLAS RESCHER
University of Pittsburgh

[20] Nachum L. Rabinovitch, *Probability and Statistical Inference in Ancient and Medieval Jewish Literature* (Toronto: University Press, 1973), pp. 127/8.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

CARTWRIGHT AND THE LYING LAWS OF PHYSICS*

One of the principal virtues of Nancy Cartwright's book, *How the Laws of Physics Lie*[1], is that it contains a sustained consideration of a feature of real science almost universally ignored by philosophers of science, namely, the use of approximations and idealizations. Typical work in the philosophy of science tends to view such elements as mere distractions, complications to be explained away in some future, more developed theory of science. One consequence of downplaying the role of approximations and idealizations is, according to Cartwright, a failure to appreciate the importance and function of phenomenological, i.e., experimental, laws. Fundamental laws, such as Newton's, get all the philosophical press. Cartwright is particularly concerned to deny the explanatory priority given to fundamental laws, what I shall call the *priority thesis*.

> A long tradition distinguishes fundamental from phenomenological laws, and favors the fundamental. Fundamental laws are true in themselves; phenomenological laws hold only on account of more fundamental ones. This view embodies an extreme realism about the fundamental laws of basic explanatory theories. Not only are they true (or would be if we had the right ones), but they are, in a sense, more true than the phenomenological laws that they explain. I urge just the reverse (100).

This paper is a close examination of Cartwright's arguments (given primarily in chapter 6 of her book) for "just the reverse." My basic contention is that these arguments, though highly suggestive, fail to take into account the relevance of the piecemeal improvability of idealizations and approximations and the corresponding improvements in predictive output. I shall develop an account of what such improvability comes to and what its positive relevance is to the priority thesis. But, as will be seen, my proposed defense will raise some new and deep problems for supporters of the priority thesis.

I. REALISM ABOUT FUNDAMENTAL LAWS

The priority thesis, as stated above, is not very precise. The natural philosophers of the seventeenth century developed creation myths, of which Cartwright gives clever caricatures, in order to give sub-

*I want to express my gratitude to the National Science Foundation (SES-8608167) and the Ohio State University for providing research funding for a larger project of which this paper forms a part.

[1] New York: Oxford, 1983. Henceforth all page references to this book will be included in the text.

0022-362X/89/8607/353–372 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

stance to the thesis (100/1). Such analyses, of course, have gone out of style and modern philosophers present the thesis in a more respectable linguistic or logical guise. Cartwright gives Adolf Grünbaum's[2] formulation as paradigmatic of what she has in mind (102).

> It is crucial to realize that while (a more comprehensive law) *G* entails (a less comprehensive law) *L* logically, thereby providing an explanation of *L*, *G* is not the 'cause' of *L*. More specifically, laws are explained *not* by showing the regularities they affirm to be products of the operation of causes but rather by recognizing their truth to be special cases of more comprehensive truths (*ibid.*, p. 14).

Cartwright calls this version of the priority thesis the "generic-specific account," since it asserts that "the fundamental laws are superior because they state the facts in a more general way so as to make claims about a variety of different circumstances as well" (103). It is worth noting, although Cartwright does not explicitly do so, the close connection between the generic-specific account and what John Earman[3] has identified as the Mill-Ramsey–Lewis analysis of law. Roughly, fundamental laws are the basic axiomatic regularities of well-confirmed or true theories. Given this status, they can be expected to contain no *ceteris paribus* or normal conditions clauses. Examples of candidates for such *ceteris paribus* free axiomatic regularities include Newton's laws, the relativistic field equations, Maxwell's equations, and the Schrödinger equation. Phenomenological or empirical laws will be those regularities derivable from the conjunction of fundamental laws with true descriptions of initial or boundary conditions. If one thinks of sound derivations as providing explanations, it follows that the phenomenological laws are explained by their sound derivations.

Now, if the generic-specific account is understood as descriptive of actual science, it is surely incorrect, since, as Cartwright observes (103), "real derivations" contain idealizations as well as approximations. Therefore, sound derivations cannot be given as explanations of phenomenological laws, since such derivations do not exist![4] Because of the falsity of the simplifications used, any derivation will be,

[2] "Science and Ideology," *The Scientific Monthly* (July 1954).

[3] *A Primer on Determinism* (Boston: Reidel, 1986), pp. 87–93. Cf. Cartwright, *op cit.*, p. 157.

[4] What are called phenomenological laws are usually only approximately true, almost always contain *ceteris paribus* or normal conditions clauses, and usually have restricted ranges of application (even when *ceteris paribus* or normal conditions are satisfied). This means that explanations of such laws may consist, in part, of unsound derivations. E.g., Boyle's law only follows from a highly simplified description of a gas. Following Cartwright, I shall ignore this complication and assume that phenomenological laws are true. An example, which I shall examine, of what Cartwright

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

strictly speaking, unsound. Therefore, confirmational strategies based on inference to the best explanation cannot be employed. So, it appears that scientific theories, sets of fundamental laws, are not testable *for their truth* by a comparison with the facts of actually derived predictions.[5] Cartwright concludes that:

> Realists [i.e., defenders of the generic-specific account] can indeed put a gloss on [actual] examples that brings them into line with realist assumptions: rigorous solutions to exact equations might possibly reproduce the correct phenomenological laws with no ambiguity "when the right equations are found." But the reason for believing in this gloss is not the practice itself, which we have been looking at, but rather the realist metaphysics, which I began by challenging. Examples of the sort we have considered here could at best be thought consistent with realist assumptions, but they do not argue for them (126/7).

There are, in effect, two challenges Cartwright directs at supporters of the explanatory priority of fundamental laws. First, to specify how actual practice can be seen as supportive of the truth of fundamental law candidates (or for that matter of their falsity). And second, to show what role, if any, the generic-specific account plays in such confirmational considerations. As a first step to meeting Cartwright's challenge, the realist will have to concede that the generic-specific account is not descriptive. Having made this concession, the next move would be to get the generic-specific account back into the picture by giving it a normative reading. Just to get things squarely on the table, let us identify this normative reading as being:

> (GS_n) The goal of science should be to seek fundamental laws which are true, and can be used in the sound derivation of phenomenological laws.

The development of derivations that satisfy the generic-specific account is now taken to be a normative ideal. Henceforth, following Cartwright, I shall mean by a *realist* (with respect to fundamental laws) one who believes that:

> (1) scientific practice can be appealed to in confirmational considerations and, in particular, that attempts to derive phenomenological laws (and more generally data) play a role in confirmational considerations;

---

has in mind is the exponential decay law. For an interesting examination of the relationship between phenomenological laws, data, and raw empirical evidence, see James Bogen and James Woodward, "Saving the Phenomena," *Philosophical Review*, XCVII, 3 (1988): 303–352.

[5] For an alternative description and discussion of this problem see my "Idealizations and the Testing of Theories by Experimentation," in *Experiment and Observation in Modern Science*, Peter Achinstein and Owen Hannaway, eds. (Cambridge: MIT, 1985), pp. 147–173.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(2) the normative version of the generic-specific account is true.

Leaving aside whatever obscurities may still be lurking here, I shall now consider two features of "typical" scientific practice which Cartwright claims are "especially damaging" for such realists.[6]

II. APPROXIMATIONS THAT IMPROVE ON LAWS

The first damaging feature is that:

> Practical approximations usually improve on the accuracy of our fundamental laws. Generally the doctored results are far more accurate than the rigorous outcomes which are strictly implied by the laws with which we begin (106).

Cartwright gives two examples of what she has in mind: the use of elementary electronics in the design of amplifier circuits, and the quantum mechanical derivation of the exponential decay law. As we shall see, these cases illustrate rather different ways that "practical approximations" can be said to "improve on the accuracy of our fundamental laws." As a consequence, the relevance of these cases for realism will be rather different as well.

I shall not repeat Cartwright's description of the amplifier case. What is illustrated is the use of an idealized, that is, simplified and hence false, description of the components in an amplifier circuit. Not surprisingly, predictions made on the basis of the combination of electromagnetic theory and these idealized descriptions are false and indeed "not even close" (109). But, since the amplifier is to some degree a modular entity, empirically determined correction factors can be systematically determined and applied which yield a new combination of theory, idealized description, and correction factor, which is predictively adequate for some practical purposes. Cases such as this are common in the history of science. A clear example is provided by Baily's early nineteenth-century experimental determination of correction factors to be applied to then existing pendulum analyses which treated air as if it operated by hydrostatic action only, i.e., which ignored viscosity effects.[7]

We now need to determine the relevance of such correction procedures for realism. Let us assume that Cartwright means to be claiming at least that:

[6] The formulation of GS_n takes its inspiration from Bas van Fraassen's similar reconstruction of the thesis of scientific realism in *The Scientific Image* (New York: Oxford, 1980), p. 8. My separation of theses (1) and (2) also mirrors van Fraassen's separation of the question of the goal of science from that of what is, in fact, achievable. But, following Cartwright, I have opted to include (1), what can be achieved, as part of realism.

[7] For details, see my "The Application of Idealized Scientific Theories to Engineering," *Synthese* (forthcoming).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

> Actual derivations will always (or nearly so) require the use of idealizations and approximations.

I shall call this the *ubiquity of idealizations* claim. The amplifier example can be understood then as an illustration of what science bound by this thesis is like (see 136/9, 154/5). Now, if the ubiquity claim is true, then it follows that actual derivations cannot (for reasons already discussed) be seen as confirming or disconfirming the fundamental laws employed.[8] Therefore, realists seem unable to give reasons based on scientific practice for thinking specific fundamental law candidates to be true or false. Their first plank is put in jeopardy by the ubiquity of idealizations claim. In response to this attack of realism, Cartwright poses the following rhetorical question on its behalf:

> But could we not "in principle" make the corrections right at the start, and write down a more accurate equation from the beginning (111)?

In order to appreciate the point of this response, one should focus on the elementary truth condition that, if a theory or set of laws is true, then applying it (validly) to a true set of initial or boundary conditions must yield a true prediction. In order to apply this condition to some fundamental law candidate, one must produce derivations that are idealization-free. But, since such soundly derived predictions are at best only possible in principle, it follows that the truth condition will be satisfied only in some *counterfactual* sense. So, for example, to achieve confirmation it must be somehow determined that, if we had the requisite computational capabilities, then the law candidates when combined with true (and sufficiently accurate but currently unknown) descriptions of initial or boundary conditions would yield correct results. For convenience and future reference, I shall interpret Cartwright's defense of realism as being based on the following truth condition:

> (R1) If a set of fundamental laws is true then it is possible in principle to construct a sound derivation of phenomenological laws, and more generally, empirical data.

But stating the requirement this way makes the situation for realists seem rather desperate. What reasons do they have for thinking individual law candidates satisfy the sort of counterfactual here envisioned? Such reasons, if available, would be confirmatory of fun-

[8] The 'or nearly so' in the ubiquity of idealizations thesis means, then, that those few derivations which are sound are insufficient for the confirmational process.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

damental law candidates. Similarly, reasons to believe such counterfactuals false would be disconfirming. But, in the absence of a specification of what such reasons will be, the truth condition R1 cannot be utilized in a defense of realism.

Actually, we need to distinguish two different but closely related issues. First, there is the question of whether R1 itself is true. This is, in fact, the issue of determinism. Perhaps it would be better to revise R1 to be a truth condition for deterministic theories.[9] Cartwright's concerns, however, do not seem to lie in this direction. When Cartwright responds to the realist defense that, "[t]his is just the assumption I challenge" (111), I believe she has in mind the need to specify how it is to be determined that *individual* law candidates satisfy the consequent of R1. That is, Cartwright's concern seems to be with the applicability and not the truth of R1. (And remember that the practical applicability of R1 is to be the basis of the defense of the first plank of realism.) The onus is on those who claim the existence of in-principle computations to give reasons for thinking this is so. From Cartwright's point of view, scientific practice does not help here, since all it does is (inductively) confirm that, in practice, sound derivations are not possible. So, the point of the amplifier example is simply to remind us that scientific practice provides no support for the in-principle derivability claims that realists seem to require.[10]

Although Cartwright's effort to isolate realist dogma is on the right track, it does contain a significant ambiguity that needs to be resolved. The 'corrections' referred to in her proposed defense of realism are not the empirically determined corrections illustrated by the amplifier and pendulum cases, but rather are meant to be either more or maximally accurate descriptions of initial or boundary conditions. In fact, this way of putting the realist's response highlights the ambiguity in Cartwright's formulation. Sometimes Cartwright seems to have the second reading in mind, i.e., that initial or boundary condition descriptions be maximally accurate (and hence true; see, e.g., p. 125). On the other hand, in her formulation of the realist response, she in fact says only that "corrections" need to be employed such that the result is "a more accurate equation." And I take it that being "more accurate" need not entail being "true" per se. This suggests the following weaker truth condition:

[9] And here one must refer to Earman's brilliant *op. cit.* The difficult question of the status of determinism in quantum physics is discussed in chapter XI.

[10] It also serves to show that practical application requires empirical tinkering.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(R2) If a set of fundamental laws is true, then we can make in principle *sufficient* corrections so as to yield better predictions.

What makes R2 interesting from the realist point of view is that piecemeal improvability of idealizations and approximations with resulting improvements in predictions is a significant feature of actual science.[11] Realists have a nice explanation of this practice of seeking increasingly more accurate and less idealized initial or boundary conditions. It begins with the observation that, since idealizations are in a sense systematically false, they introduce a bias or distorting factor into the derivation of predictions. So, it follows that fundamental laws, even if true, will produce false and distorted predictions. Only in the highly unlikely case where there are exactly compensating simplifications will predictions come out true or accurate to within experimental error.[12] From this perspective, the aim of science should be to construct more accurate models where the anticipation is that our theories, if true, will produce more accurate predictions when applied to these more accurate models. For example, in the pendulum case, Baily's correction coefficients were rendered superfluous and corrected by Stokes's development of a Newtonian theory of viscous fluids. Similar cases can be found in the field of amplifier design. Cartwright herself, perhaps inadvertently, gives several such examples from quantum mechanics.[13] The relevance of the existence of such improvability is that it is inductive evidence in favor of the satisfaction of the consequent of R2, a necessary condition for truth. Similarly, failure to generate improved predictions on the basis of improved idealizations and approximations (as input to theory) is evidence against the satisfaction of the consequent of R2, and hence for the falsity of the theory in question. What this suggests then is the following pair of confirmation principles:

(C1) A set of fundamental laws receives confirmation if the use of more realistic specifications of initial or boundary conditions in fact leads to more accurate predictions.

[11] Supporting cases (including the Michelson-Morley and Kennedy-Thorndike experiments, Brownian motion, and Newton's optics and gravitation theory) are discussed (and further references given) in my "Idealizations and the Testing of Theories by Experimentation."

[12] For an extremely interesting study of this sort of bias in biological explanations, see William C. Wimsatt's pioneering "Reductionistic Research Strategies and Their Biases in the Units of Selection Controversy," in *Scientific Discovery, vol. II: Historical and Scientific Case Studies*, Thomas Nickles, ed. (Boston: Reidel, 1980), pp. 213–259.

[13] See her descriptions of Messiah's Hydrogen Atom (137/8) and Louisell's treatment of the gas laser (146–148).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(C2) A set of fundamental laws receives disconfirmation if the use of more realistic specifications of initial or boundary conditions does not lead to more accurate predictions.

The idea then is that acceptable theories, i.e., sets of fundamental laws, be *monotonic toward the truth* in the sense that more accurate and less idealized initial condition descriptions lead to more accurate predictions. The proposal encapsulated in C1 and C2 is that confirmation be seen as an induction of the sort: from theory T is monotonic in known instances (and other relevant knowledge), conclude that T is, in principle, monotonic, that is, that it satisfies the consequent of R2.[14] (I discuss below the problem of getting from the consequent to the antecedent of R2.) If correct, these theses show how fundamental law candidates can be confirmed (or disconfirmed) even in the face of the use of distorting approximations and idealizations. So, the first plank of realism can be secured against Cartwright's challenge by C1 and C2. Now, if all of this is correct, there is strong support as well for the second plank, GS_n, the normative version of the generic-specific account. For given the confirmation theses C1 and C2, it follows immediately that we should aim for increasingly more accurate derivations of phenomenological laws and individual predictions, since in that way we confirm and disconfirm our law candidates. To be sure, thesis GS_n demands that we seek out a goal that is likely to be at best an in-principle possibility only. Cartwright can perhaps be read as suggesting that the practical unobtainability of the goal is a reason for rejecting it. But, from the perspective of the confirmation principles C1 and C2, this would be a poor reason for rejection, since the real issue is the continuing piecemeal improvability of our models and calculations. The idea is that only by seeking maximally accurate derivations will we generate the piecemeal improvements that are sufficient to support the proposed confirmational practice.

There is a complication in the application of C1 and C2 which should now be mentioned. Typically, particular instances of improvement or nonimprovement (of prediction due to more realistic input) will not be decisive with respect to the satisfaction of the consequent of R2, since improvability may ultimately come to an end and recalcitrant cases may ultimately yield better predictions given increasingly more realistic input. What this means at the least is that C1 and C2 must be interpreted as being in some sense prima facie

[14] Variants of C1 and C2 can be found in the work of Thomas Kuhn and Imre Lakatos, as well as in Pierre Duhem.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

inductive confirmation principles. In particular, the computational specifics (how the theory is computationally implemented) may override the prima facie confirmational value of existing samples of improvement or nonimprovement. This complication will be further examined in the final section when I discuss Cartwright's second line of argumentation against realism.

Before proceeding, I want to register the claim that my proposed confirmation principles C1 and C2 are, in fact, employed in science. I have in other published work given several illustrations of this practice.[15] So, not only can the actual scientific practice of generating increasingly more realistic derivations be used to confirm or disconfirm fundamental law candidates, the confirmational procedure suggested here (based on C1 and C2) is, in fact, that adopted by actual science. From my point of view, scientific practice is as it ought to be. Especially noteworthy are those cases where impossibility proofs (of varying rigor) were given to show that it is, in principle, impossible to derive some value or function from a set of proposed fundamental laws and initial condition or boundary descriptions, however accurate. Realists have a ready explanation of the importance given to such cases in the history of science. If the target prediction is true then, since the consequent of R2, as well as that of R1, is not satisfiable, then the fundamental law candidate stands refuted. Classic examples of such cases include kinetic theory and specific heats, aether theory and the null result of the Michelson-Morley experiment, and the singularities entailed by the general theory of relativity. (Newton's derivation of Kepler's laws has closely related properties and will be discussed below.) In the case of specific heats, it was provable (or at least believed so) that the equipartition theorem entailed that making molecule descriptions more realistic (by adding more degrees of freedom) would lead to predicted values for specific heats which would be progressively worse than those predicted on the basis of simple models.[16] In the Michelson-Morley case, proponents of the aether theory suggested that more realistic analyses of the experiment would yield predictions consistent with a Fourier decomposition of the Michelson-Morley data. Lorentz, however, was able to give a general proof, based on the calculus of variations, showing that more realistic analyses would not yield pre-

[15] See fn. 11.

[16] An excellent review is given in Peter Clark, "Atomism versus Thermodynamics," in *Method and Appraisal in the Physical Sciences*, Colin Howson, ed. (New York: Cambridge, 1976), pp. 41–106.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

dictions consistent with Michelson and Morley's data.[17] Finally, in the relativistic singularities case, it was originally thought that the singularities were merely mathematical debris generated by the excessively symmetrical and simple conditions assumed in the initial solutions of the field equations. Intensive mathematical investigation revealed, however, that the singularities were an abiding shortcoming of the field equations.[18] The importance of cases such as these is exactly explained by realism, since in these cases it could be shown that the goal set by the normative version of the generic-specific account is not obtainable even in principle. Hence, such theories were either rejected or seen as, at best, stepping stones to more adequate theories.

My conclusion, then, is that the principles C1 and C2, along with R2, are sufficient for establishing realism (as defined above). The stronger R1 is not necessary. There does seem, however, to be some connection between R1 and R2. I claimed above that C1 and C2 derive their plausibility from R2. But, if we now ask what justifies R2, we run into some difficult questions. One possible realist strategy would be to claim that there exists some continuity condition such that R1 and R2 are equivalent given that condition. That is, that theories should be continuous in the sense that monotonic improvability yields true output in the limit. If so, then C1, by providing inductive support for the consequent of R2, indirectly supports satisfaction of the consequent of R1 as well; similarly for C2 and the nonsatisfaction of the consequents in question. This suggests a realism based on R1, some continuity condition, and some variant of determinism. Since the examination of how such a position might be developed and supported would require a significant expansion of my methods and focus, I shall rest content in this paper with having run the realist down to this level.[19] To summarize: my claim is that

[17] Some of the details are discussed in my "The Michelson-Morley Experiment and the Appraisal of Theories," in *Scrutinizing Science: Empirical Studies of Scientific Change*, Larry Laudan, Rachel Laudan, and Arthur Donovan, eds. (Boston: Kluwer, 1988), pp. 245–266.

[18] This history is comprehensively reviewed in F. J. Tipler, C. J. S. Clarke, and G. F. R. Ellis, "Singularities and Horizons—A Review Article," *General Relativity and Gravitation: One Hundred Years after the Birth of Albert Einstein*, A. Held, ed. (New York: Plenum, 1980), pp. 97–205.

[19] For the sake of completeness, I should also note that there is a problem in justifying the move from satisfaction of the consequents of R2 and R1 to their antecedents. One response would be simply to deny that such movement is possible, that is, to claim that all there is to confirmation is the satisfaction or nonsatisfaction of necessary conditions for truth. At the other end of the spectrum is the equally natural move of defining the truth of *quantitative* theories as being the conjunc-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Cartwright ignores the confirmational value of the sort of piecemeal improvability of idealization and approximation which is common in actual scientific practice. If we accept such improvability as inductive support for the satisfaction of the consequent of R2, then Cartwright's challenges to realism are answered. This approach, however, generates a new problem for realism, namely, to justify this type of induction. So, although I may disagree with Cartwright about some of the specifics of her argumentation, that argumentation disambiguated and carried to its natural conclusion does isolate a significant chunk of realist dogma.

The reader will perhaps remember that Cartwright gives two examples in support of her thesis that "practical approximations usually improve on the accuracy of our fundamental laws." I now turn to an examination of her second example. Cartwright describes two methods of deriving the exponential decay law from quantum mechanics: the classic Weisskopf-Wigner approach and the more recent Markov-type treatment. In both cases, approximations and idealizations are used. But, despite the use of such simplifying assumptions, the exponential decay law is derivable exactly. So, from the point of view of the above discussion, realists must interpret this as a fortunate cancellation of the biasing effects of idealization and approximation (cf. 140). But even more interesting is the fact that, given the nature of the approximations and idealizations used, it is clear that, if more exact methods were used, the predictions obtained would, in fact, differ from the exponential decay law. To take into account the small terms and effects ignored would displace the prediction from that obtained on the basis of existing approximations and idealizations. As summarized by Cartwright, the exponential decay law "cannot be derived exactly in the quantum theory. [It] can only be derived by making some significant approximation" (113). The lesson to be drawn from this case is, according to Cartwright, that:

> The fact remains that the data, together with any reasonable criterion of simplicity (and some such criterion must be assumed if we are to generalize from data to laws at all) speak for the truth of an exponential law; but such a law cannot be derived rigorously. Thus it seems that the approximations we make in the derivation take us closer to, not further from, the truth (113).

---

tion of the consequents of R1 and R2. For more discussion and consideration of intermediary positions, see the references given in fn. 25.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

My first comment about this case is to note that it is rather different from that of the amplifier case. The latter is merely an example of how, given an idealized analysis, empirically determined correction factors can be incorporated to create a computation with some predictive and engineering value. Improvability of analysis is not shown impossible merely by the fact that correction factors are used at some early stage of investigation (in fact, Cartwright seems to admit this on p. 109). The exponential decay law case, however, is one where (1) no empirically determined corrections are required if appropriate approximations and idealizations are used, and (2) improvements in the realism of the analysis will lead to divergence from the desired prediction. This latter feature would seem to entail that the fundamental equations of quantum mechanics are false, since those equations plus true initial or boundary condition descriptions must yield some law that is contrary to exponential decay. Somewhat surprisingly, Cartwright does not push for this result, but instead makes only the weaker claim that "the approximations we make in the derivation take us closer to, not further from, the truth." Of course, as Cartwright later notes, it is currently impossible to test the accuracy of the exponential decay law with the precision needed to test it *vis-à-vis* predictions made on the basis of more accurate quantum calculations or models. This being so, the proposed disconfirmation of the fundamental laws of quantum mechanics would require that greater weight be given to some simplicity principle (as applied to our data) than the prediction of an otherwise well-confirmed theory. The bottom line then is, or should be, that this case cannot be used to show the Schrödinger equation false, since experimental data are not available to sufficient precision. But, if this is correct, then it also follows that Cartwright's weaker claim (i.e., "that the approximations we make in the derivation take us closer to, not further from, the truth") ought also to be rejected.

There are some interesting similarities here with Newton's derivation of Kepler's laws. Even more clearly than in the case of the exponential decay law, Newton's target predictions, Kepler's laws, follow only if certain idealizations and approximations are used: that there are only two bodies, and that only one of the bodies acts on the other. If more accurate descriptions of the solar system are used, then, because of perturbations, Kepler's laws will not follow. Newton saw this as a good test of his theory and directed Flamsteed to look for the requisite perturbations. Unlike the quantum mechanics case, the perturbations were detectable given then available means of observation. This case provides a nice example of my proposed confirmation principles C1 and C2 in action, since more realistic analyses

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

of planetary motion utilizing perturbation could be expected to accommodate or approach the observationally determined variations from Kepler's laws.[20]

Cartwright's presentation of the exponential decay case contains an important ambiguity. When she says, for example, that the exponential decay law "cannot be rigorously derived," the most natural reading is that an in-principle calculation using true initial conditions would not yield the decay law. On this reading, improvements in the accuracy of the initial conditions, no matter how finely made, will never lead to the satisfaction of confirmation principle C1. In other words, disconfirmations made on the basis of C2 cannot be viewed as temporary phenomena and mere mathematical artifacts. But I do not think that Cartwright must be read in the strong sense just suggested. The details of the case suggest a weaker reading. For example, both the Weisskopf-Wigner and the Markov approaches begin with what appear to be idealizations or simplifications. For example, the Weisskopf-Wigner method begins with the assumption "that the only significant coupling is between excited and the de-excited states" (115). If this is so, then what we have is that improvements made against a backdrop of initially made simplifications will yield predictions that diverge from the truth. Cartwright can be read then as asserting that, given the current calculational basis, improvements will cause a divergence from the exponential decay law.

What is the relevance, given its various readings, of the exponential decay case to realism? If decay could be determined to be truly exponential, and if my first interpretation of Cartwright were the correct way to view the case (i.e., all improvements will lead to a worsening of predictions), this would be at best only sufficient to refute quantum mechanics. I do not see how this refutation would affect realism. If anything, such a refutation would serve to reinforce realism, since showing the in-principle impossibility of soundly deriving the exponential decay law provides, from the point of view of realism, the most compelling disconfirmation possible. In fact, I would speculate that Cartwright avoided using the exponential decay law to disconfirm the Schrödinger equation precisely because to do so would be to adopt a testing methodology so strongly in accordance with realism.

What happens if, as seems sensible, we do not interpret exponential decay as a refutation of quantum mechanics? As indicated, there

[20] For more details, see my "Newton's Demonstration of Universal Gravitation and Philosophical Theories of Confirmation," in *Minnesota Studies in the Philosophy of Science* X, J. Earman, ed. (Minneapolis: Minnesota UP, 1983), pp. 179–199.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

are two possible ways to do this. In the first, we interpret the case as a standoff between fundamental and phenomenological law, because the requisite data are not available with sufficient accuracy. On this approach, we have what seems to me to be a conscientious attempt to apply C1 and C2 which is stymied by experimental limitations. If this were always the case, then our confirmational attempts could never get off the ground and the first plank of realism would be false. But the exponential decay case is not, I contend, typical of scientific practice. A second way to view the exponential decay case as not disconfirming quantum mechanics would be to claim that, even if experimental data were available with requisite accuracy, C1 and C2 could not be utilized, since there is an unavoidable and ineliminable set of basic idealizations common to any actual calculation of the decay law. (This approach is based on my second reading of Cartwright.) The idea here is that the existence of such a set means that the presence or absence of actual monotonic behavior is not a reliable indicator of long range or in-principle behavior.[21] Although this approach saves quantum mechanics from refutation, it has nasty consequences for realism. For if the existence of unavoidable and ineliminable sets of basic idealizations is common to all scientific calculations, then the proposed confirmational procedure based on C1 and C2 cannot be utilized. Hence, realists lose support for their first plank and motivation for their second (GS_n). Exactly why actual monotonic behavior in the sort of case considered might not be a reliable indicator of long-range behavior needs to be examined. But, before further discussing this potential embarrassment for realism, I want first to consider Cartwright's second line of argumentation against realism. As will be seen, that argumentation will suggest similar difficulties for realism.

III. APPROXIMATIONS NOT DICTATED BY THE FACTS

Cartwright's second line of attack on realism is based on the observation that "the choice [of approximations and their application] is constrained, but not dictated by the facts [about initial or boundary conditions], and different choices give rise to different, incompatible results" (107). If this is the case, then the "generic-specific account fails because the content of the phenomenological laws we derive is not contained in the fundamental laws which explain them" (107). Cartwright gives two examples of what she has in mind: Bethe's derivation of the Lamb shift (in the excited state) by emendation of the Weisskopf–Wigner method of treating exponential decay, and Agarwal's derivation of the Lamb shift in the ground state. Since

[21] In fact, Cartwright gives a variant of this line of argumentation in ch. 8, *op cit.*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by ANTSANE Foundation Chennai and eGangotri

Cartwright sees similar strategies employed in these cases, I shall restrict attention to the first. In the Weisskopf-Wigner method, a sum of values over the modes of the field is replaced by an integral. Since the modes form a near continuum, this replacement can be expected to cause a relatively small error. In addition, terms that vary slowly with frequency are treated as constant and hence factored out. I shall refer to this replacement of sum by integral and associated factoring of terms as the *sum approximation.* Another approximation used involves factoring out from a time integral a term that varies relatively slowly and then extending the limit of the integral to infinity. Again the physical properties of the atom-field pair suggest that the error introduced by such an approximation will be relatively small. I shall refer to this approximation as the *time approximation.* In the original Weisskopf-Wigner calculation of the exponential decay law, the sum approximation was applied before the time approximation. This order made for a more convenient calculation, since certain terms not relevant for decay, and hence mere "mathematical debris," were eliminated from consideration. As Bethe realized, however, these eliminated terms could be interpreted once the discovery of the Lamb shift indicated that something more was needed. What Bethe did was to reverse the order of application of the time and sum approximations and show, firstly, that exponential decay could still be derived and, secondly, that what was formerly considered to be merely mathematical debris could be utilized to account for the Lamb shift. Since Bethe's calculation occurred in response to the discovery of the Lamb shift, it was somewhat ad hoc. Cartwright claims that what Bethe's procedure shows is that "we do not derive [the Lamb shift] from [the fundamental quantum equations]" (120). This is surely correct, if understood to be merely claiming that Bethe's derivation was not sound, that is, that idealizations and approximations were still used. But Cartwright means to be getting at more than just this. To see what she is getting at, begin by observing that Bethe did not show that a calculation using maximally precise values which was also free of all idealizations and approximations would still lead to the Lamb shift (cf. 125/6). In other words, Bethe did not demonstrate that quantum mechanics satisfies the consequent of R1 with respect to the Lamb shift. Therefore, all that had been shown, and after the fact to boot, is that some combination of approximation and idealization entails the Lamb shift. But, since there also exists some combination that does not entail the shift, it follows that whether or not quantum mechanics entails the Lamb shift depends on the idealizations and approximations that *we decide to use.* Hence, according to Cartwright, the "generic-specific ac-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

count fails because the content of the phenomenological laws we derive is not contained in the fundamental laws which explain them" (107).

There is an important objection to be made to this line of argumentation, since Bethe did more than just show that some combination of idealization and approximation could be used to derive the Lamb shift. As Cartwright's own description of the case shows, it is demonstrable that the order of application of approximations does affect whether or not certain terms remain in the calculation. So, in this case, one order is clearly preferable, and hence more realistic, because less information is lost (see pp. 120, 122). Therefore, Bethe's modification would seem to illustrate a successful application of the confirmation principle C1 and hence pose no threat to realism. I do not think Cartwright would be happy with this line of defense. I claim that Bethe's reordering of the application of approximations is superior, because it preserves terms eliminated in the original Weisskopf–Wigner method. But Cartwright would object to this appraisal, since for her the preservation of terms cannot be a general principle for awarding (before experimental testing) the status of being more realistic to one set of idealizations over another. This is because, as she puts it (in a slightly different context), such a principle "assumes 'the more accuracy, the better'. But this is frequently not so . . ." (121). The idea is this. Sometimes what seem to be more complex and hence more realistic accounts generate what turn out to be uninterpreted and uninterpretable mathematical debris. If this is so, then one cannot judge before experimental testing that the generation of extra terms makes one account more realistic than another. Therefore, one cannot employ my proposed confirmation principles C1 and C2, since they require a prior determination of the relative realism of competing idealized and approximate treatments. Bethe's realization that the mathematical debris generated when the time approximation was applied before the sum approximation could be interpreted as the Lamb shift may have served as a good reason to grant his calculation the status of being more realistic than the original Weisskopf-Wigner calculation. This is to use experimentation to test the relative realism *of the idealizations,* however. It is not to use the previously determined relative realism of idealizations to test fundamental theory! So, the Bethe case does not illustrate a successful application of the confirmation policy encapsulated in principles C1 and C2.

There are two responses to be made to this line of attack. First, the type of difficulty exemplified by the Bethe case does not infect all applications of quantum mechanics or other scientific theories. Se-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

cure judgments of relative realism (made against a relatively stable set of background assumptions) are common. In fact, as already noted, Cartwright herself makes several such judgments throughout her book. So, the problem illustrated by the Bethe case is a local and not a global problem. Second, even if judgments of relative realism were *never* available before the accuracy of predictions was compared, the situation for the applicability of C1 and C2 would still not be hopeless. In this paper, I can only briefly sketch what I have in mind.

Let T be the law or conjunction of laws whose truth we wish to ascertain. And assume that I_1 and I_2 are two competing combinations of idealization and approximation whose relative realism cannot be determined according to background standards in place. Assume, finally, that $T \,\&\, I_1 \Rightarrow P_1$, and $T \,\&\, I_2 \Rightarrow P_2$, and that experiment reveals that prediction P_2 is more accurate than P_1. In such a case, we would (as shown by Cartwright's own cases) claim (perhaps only tentatively) that I_2 is, for the situation considered, call it s, the more accurate description. Now, if new cases of theory application can be judged relatively similar to s, then we should expect I_2 to continue to be superior to I_1. More accurately stated, those features of I_2 which carry over to the new calculation should *ceteris paribus* remain relatively more realistic than the corresponding features of I_1 which are carried over to the new cases. So, new calculations based on I_2 should continue to produce better results than ones based on I_1. For example, if a hydrostatic as opposed to a vacuum treatment of the medium of pendulum motion yields better predictions, we should expect a hydrostatic treatment to continue its superiority when applied to compound pendulums. What this means in terms of C1 and C2 is clear. In cases where the relative realism of two (or more) treatments or models cannot be determined, one should perform an experiment to determine in an ad hoc way the relative realism of the competing treatments. Once such a judgment is in place, the set of underlying fundamental laws can be tested by applying it in new cases where I_1 and I_2, or relevantly similar variations, can be applied. On the view I am proposing, one seeks to maximize the coherency of (1) judgments of the relative realism of competing combinations of idealization and approximation, with (2) judgments of the improvability of predictions of a theory that is to be tested. It should be noted, however, that it is an empirical possibility that the complexity of phenomena and computation will overwhelm our efforts to maximize the sort of coherency proposed. If such maximization cannot be carried out, then the confirmation principles C1 and C2 cannot get off the ground. Some of Cartwright's remarks can be interpreted as ex-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

pressing support for the position that such overwhelming complexity is the case (see, e.g., 145). But how is the existence of overwhelming complexity to be determined other than by adopting the generic-specific account as a normative standard? So, once again the complications exemplified by Cartwright's examples serve, or so I claim, to strengthen the reasonableness of the second plank of realism.

Cartwright has another objection to the principle "the more accuracy, the better." It is that "the initial problem is set only to a given level of accuracy, and any accuracy in the conclusion beyond this level is spurious" (121). It seems to me that, although this consideration may cut against the principle if used in engineering contexts, it does not so cut when the principle is used in confirmational contexts. It would be absurd, for example, to use relativistic considerations when designing a bridge or, for that matter, when sending men to the moon. On the other hand, if, in a confirmational context, measurement procedures are not available to accuracy sufficient to test the predictions obtained by competing idealizations, this serves as a justification for experimentalists to improve their procedures. The history of science is not lacking in cases when experimentalists developed their techniques in response to calculational advances.[22]

Having defended the principle "the more accuracy, the better" against Cartwright's criticisms, I now want to conclude this paper by briefly raising an objection of my own to the principle. It does seem possible, because of the computational particulars, that, even if some idealization really is more realistic than another, the prediction obtained on the basis of the superior idealization may nevertheless not turn out more accurate. If this is so, then use of principle C2 would render the theory in question disconfirmed, which would be to blame the theory for what seems to be merely a computational quirk. This potential problem with the application of C2 (and *mutatis mutandis* C1) is, I believe, a close relative of the problem with which I closed the preceding section (viz., that the existence of an enduring set of idealizations means that the presence or absence of actual monotonic behavior may not be a reliable indicator of long-range monotonic behavior). A simple abstract example will serve to motivate consideration of the possibilities in question.

Assume that the true laws for some set of phenomena are:

[22] In fairness to Cartwright, I should say that she probably had in mind the exponential decay law case where the desired prediction was obtainable on the basis of an idealized and approximate treatment. But the relevance of this sort of case for realism has already been discussed.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

$$0.0001x_1 + x_2 = 1$$

$$0.5x_1 + 0.5x_2 = 1$$

Consider now three scientists who all believe the correct laws to be of the form:

$$\alpha x_1 + x_2 = 1$$

$$0.5x_1 + 0.5x_2 = 1$$

The first scientist, however, believes α to be zero. As a result, his computational task is quite easy and yields $x_1 = 1$ and $x_2 = 1$. The second and third scientists correctly believe α to be .0001, and both solve the equations by Gaussian elimination beginning with variables of lowest coefficient. All the scientists use floating-point arithmetic, with the first two scientists being restricted to three-digit precision, while the third has access to five digit precision. Call the theories associated with each of these scientists T_1, T_2, and T_3, where I mean to include the associated calculational procedures as part of the respective theories. Now, it would appear that in terms of relative realism, based on the principle "the more accuracy, the better," theory T_1 is less realistic than T_2, and T_2 is less realistic than T_3. We would expect, therefore, that the predictions of these theories be similarly monotonic, i.e., that x_1 and x_2 calculated by T_2 will be more accurate than values calculated by T_1, and that x_1 and x_2 calculated by T_3 will be more accurate than values calculated by T_2. This, somewhat surprisingly, is not the case, since

$$T_1 \text{ yields } \langle x_1, x_2 \rangle = \langle 1.00, 1.00 \rangle;$$

$$T_2 \text{ yields } \langle x_1, x_2 \rangle = \langle 1.00, 0.00 \rangle; \text{ and}$$

$$T_3 \text{ yields } \langle x_1, x_2 \rangle = \langle 1.0000, .99990 \rangle.$$

So, in the case of T_2, the more accurate value for α actually leads to a worsening of prediction, given the limited three-place computational accuracy and the fact that Gaussian elimination based on least coefficients is used.[23]

Now, one might be inclined to argue that this case does not even suggest that Cartwright is correct in her criticism of the claim that the more accuracy the better, since, as the example shows, more

[23] The second scientist first solves for x_1 and obtains (using his floating-point arithmetic) $x_1 = 10{,}000 - 10{,}000x_2$. Substituting in the second equation yields: $5{,}000x_2 = 5{,}000$. So $x_2 = 1$. Back substitution (in the first equation) yields $x_1 = 0$. For details on the sort of accuracy problem illustrated here, see Vasek Chvátal, *Linear Programming* (New York: Freeman, 1980), pp. 74–77.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

accuracy does ultimately win out. This response misses, I contend, the proper lesson to be drawn from this (admittedly simple) example. My basic thesis has been that realism can be supported, if one accepts the principle that samples of monotonic behavior serve as prima facie inductive evidence for the in-principle existence of sound derivations. But what the current example suggests is that a theory may be retained in the face of nonmonotonic behavior on the ground that computational quirks are involved. Ultimately, more accuracy may win out, but, in the absence of known-true theories (and the requisite computational capabilities), we will not know when this will occur.[24] If this is so, then C1 and C2 seem of negligible applicability. And, if this is so, then Cartwright's case is made against realism. Given the ubiquity of idealizations and approximations, standard confirmational procedures cannot get off the ground. Defeating this result with authority or embracing it with substantiated enthusiasm requires a deeper analysis of the computational nature of science than achieved in Cartwright's book or in this paper.[25] But Cartwright is to be greatly praised for having moved us in the direction of such deeper analysis in so elegant and suggestive a way.

RONALD LAYMON

Ohio State University

[24] Some real cases are briefly analyzed in my "Idealizations and the Testing of Theories by Experimentation," p. 158, as well as my "Newton's Demonstration of Universal Gravitation and Philosophical Theories of Confirmation," pp. 190/1.

[25] Proposals for this deeper analysis are contained in my "The Confirmational and Computational Differences between the Social and the Physical Sciences," *The Israel Colloquium: Studies in History, Philosophy and Sociology of Science,* vol. VI, Edna Ullmann-Margalit, ed. (Boston: Kluwer, forthcoming); and my "Using Scott Domains to Explicate the Notions of Approximate and Idealized Data," *Philosophy of Science,* LIV, 2 (1987): 194–221.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE VIRTUES OF IGNORANCE*

> If anyone will sum us up by our actions and conduct, a greater number of excellent men will be found among the ignorant than among the learned: I mean in every sort of virtue. The old Rome seems to me to have borne men of greater worth, both for peace and for war, than that learned Rome that ruined itself. Even if the rest were exactly equal, at least worth and innocence would remain on the side of the old, for they dwell singularly well with simplicity.
>
> —Montaigne, "Apology for Raymond Sebond"[1]

THE study of virtue and character has recently received great attention. One common element in almost all accounts of the virtues is the consideration that, for x to be a virtue, the agent possessing x exercises it knowingly. On an Aristotelian view, for example, virtues involve practical wisdom. According to the doctrine of the mean, the person of practical wisdom chooses the mean disposition that lies between a deficiency and an excess. The virtuous man is the one who avoids excess and defect by conforming to the mean with respect to his actions and passions. The virtuous act is the act performed in accordance with the mean disposition. On Aristotle's account, the agent must know what he is doing in being virtuous, because he is making a choice based upon practical wisdom. That is, in order for a person to act according to virtue, that person must know what he is doing under the description that makes the action characteristic of the virtue. A consequence of such a view of virtue is:

(A) No virtue is constituted by, or based upon, ignorance.

For example, it cannot be the case that the generous person, in exercising his generosity, does so unknowingly. For a person to be generous to the needy, that person must pick up on the need of others, and respond to that need by volunteering aid. Thus, (A) is necessary to most accounts of virtue, because on these accounts being virtuous is largely a matter of knowing what to do. But I shall

* This paper has benefited from conversations with many people. I would particularly like to thank Douglas Butler, Don Hubin, Thomas Nagel, Wendell O'Brien, J. B. Schneewind, Walter Sinnott–Armstrong, Michael Slote, Roy Sorensen, and Susan Wolf. I would also like to thank the audiences of colloquia at Carnegie-Mellon, Georgetown, and Virginia Tech.

[1] *The Complete Essays of Montaigne*, Donald M. Frame, trans. (Stanford: University Press, 1976), p. 359.

0022-362X/89/8607/373–384 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

be arguing against (A) in this essay by discussing a class of virtues which involve ignorance in an essential way and which I am calling the *virtues of ignorance*.[2]

I shall take modesty as my paradigm case of a virtue of ignorance. The first two sections of this essay will deal with an analysis of modesty along these lines. I will be taking time and care with this analysis, since I feel sure that many will question an analysis of modesty as a character trait that requires ignorance. Other examples of virtues of ignorance will also be discussed, and I shall close the essay with a general discussion of them.

I.

> 'My dear Watson, . . . I cannot agree with those who rank modesty among the virtues. To the logician all things should be seen exactly as they are, and to underestimate one's self is as much a departure from truth as to exaggerate one's own powers'.
>
> —Sherlock Holmes, from "The Greek Interpreter" by Sir Arthur Conan Doyle

'Modesty' has at least two senses—there is the sexual sense of modesty, usually considered a womanly virtue, which primarily consists in a chaste and unassertive countenance. There is also the more usual sense that is associated with self-deprecation, or an underestimation of one's self-worth. It is this latter sense which concerns me in this essay. The above quotation encapsulates a good deal of what I want to say about this sense of 'modesty', for I largely agree with what Holmes has to say about modesty. A modest person *underestimates* self-worth. It is this claim which I shall be arguing for in the first part of this paper. Modesty is dependent upon the epistemic defect of not knowing one's own worth. Unlike Holmes, however, I do believe that modesty is a virtue. Modesty, as a virtue, therefore undermines (A).

I anticipate that this statement of mine on the nature of modesty will cause some readers to protest. It is too narrow, they will claim, because we want to count as modest those people who merely understate their self-worth, though they are fully aware of their own self-worth. Such people exhibit false modesty, perhaps, but not sincere modesty. In any case, to claim that my account of modesty is too

[2] One possible example of a virtue of ignorance was discussed by Herbert Morris in his essay "Lost Innocence," in *On Guilt and Innocence* (Berkeley: California UP, 1976). In that essay, Morris presented an analysis of innocence which treated the phenomenon of losing innocence as akin to the getting of wisdom. That is, innocence itself is a sort of ignorance. In the case of Adam and Eve, the case on which Morris focuses his discussion, the loss of innocence is the result of the acquisition of knowledge of good and evil.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

narrow will not harm the larger point that I wish to make—that modesty is a virtue that undermines (A). For it is granted by all that there are some cases of modesty where the person does underestimate self-worth, i.e., where the person lacks an accurate understanding of himself. At least some cases of modesty involve the epistemic defect I refer to above. Before going into this issue in more detail, I would first like to discuss several alternative analyses of modesty.

A desired feature of any account of modesty is that it explain the oddity of:

(1) I am modest.

(1) seems to be oddly self-defeating. If I were to utter (1), the charitable person would think that I was joking. Others would think that I was being nonsensical. Each of the three accounts of modesty which I shall be considering will be able to analyze (1), but two will be ruled out due to other problems.

There are three accounts of modesty which seem to be initially plausible. The first one that I shall discuss is the *behavioral account.* On this analysis, modesty is equated with some sort of behavior, such as the careful avoidance of boastfulness. This account could explain the oddity of (1), since in uttering (1) the utterer would be bragging. Therefore, his behavior would serve as evidence against the truth of the utterance—as is the case, for example, of someone saying "I don't talk." So the behavioral account is able to handle this particular desideratum.

If, however, the behavioral account were correct, propositions such as

(2) Ron does not brag, but Ron is not modest.

would be contradictions, since not to brag *is* to be modest. But (2) makes sense, and this would be the case with any sort of behavior one is inclined to equate with modesty. Consider, for example, the following story: Robinson Crusoe has just been marooned on his deserted island. He builds himself a house, grows his own food, makes his own clothes, and, in the end, develops a very high, though accurate, opinion of himself. Yet he does not display any "immodest" behavior, such as bragging. Nevertheless, he is immodest, because he still retains a very high opinion of himself—he thinks he is just wonderful. One possible objection to this example is that we could say that Robinson Crusoe has a disposition to brag, but, since no one is on the island with him yet, he merely lacks the opportunity. But it is not difficult to imagine a man who, though immodest, is not disposed

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

to brag, or to exhibit immodest behavior—he may, for instance, think such behavior in bad taste, while still retaining an inflated view of his self-worth.

The second account of modesty I shall call the *understatement account*. This account is actually a variety of the first. On this account, a modest person knowingly understates his self-worth, perhaps because he wishes to avoid being a braggart. This analysis can also explain the oddity of (1). By uttering (1) while knowing himself to be modest, the agent would not be understating, and thus not modest. For this reason, the expression of (1) is odd. This account, however, is vulnerable to a serious objection: the analysis is an account of false modesty, rather than of genuine modesty. If a person should understate her self-worth, and others had reason to believe that she knows herself to be better than she says she is, they would tend to call this *false* modesty on her part.

I favor a third analysis that I have chosen to call the *underestimation account of modesty*. On this account, the modest person underestimates his self-worth. If he speaks, then he understates the truth, but he does so unknowingly. This entails that the modest person is ignorant, to a certain degree, with regard to his own self-worth.[3] He underrates himself, and therefore only takes a portion of the credit due him. It is important to note here that the modest person must still think his character and deeds to be of some worth. For instance, the person who thinks he has done something wicked when, in fact, he has done something good is not modest. In this way, modesty differs from complete self-deprecation. This account seems plausible to me, because sincerity seems to be a necessary condition for genuine modesty, and the person who underestimates his self-worth is sincere in understating it.

This account can also explain the oddity of (1). Modesty involves a lack of knowledge regarding self-worth. In general, when one asserts *p*, one is suggesting that one knows, or at least believes, *p*. In uttering (1), I imply a realization of my lack of knowledge with regard to modesty, making it no lack of knowledge. It is self-defeating to utter (1). Like Moore's "It is raining, but I do not believe it," (1) can be true, but I cannot believe it to be true or utter the sentence sensibly. I can be modest, but I cannot *know* it.

I should like to underscore at this point the difference between this underestimation analysis, and the understatement analysis. Con-

[3] It need not be his own self-worth, at least directly. A person could be modest with respect to his or her children and their accomplishments. Even this sort of modesty, however, involves self-worth, if one believes that the deeds of one's children reflect on oneself. And, for sincere modesty, it is still a matter of underestimation.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

fusing genuine modesty with understatement is like confusing lying with telling a falsehood. Genuine modesty does not always involve understatement, and understatement is not always a case of genuine modesty. The two are very closely related, and it may often be the case that, where there is understatement, there is genuine modesty —just as it may often be the case that, where there is a falsehood being told, someone is lying. But, just as lying should not be equated with telling a falsehood, genuine modesty should not be equated with understatement. Genuine modesty is a matter of underestimation, which often results in understating self-worth.

What the analysis comes down to is this: for a person to be modest, she must be ignorant with regard to her own self-worth. She must think herself less deserving, or less worthy, than she actually is. There may be cases of modesty which involve a person simply not thinking of himself at all. But these cases, even though they do not involve underestimation, do involve ignorance. If the person has no opinion of his self-worth, then he is ignorant of it. Since modesty is generally considered to be a virtue, it would seem that this virtue rests upon an epistemic defect.[4]

It could be objected that this analysis of modesty is too narrow, that there are cases of modesty where the person is aware of his self-worth but simply does not want to talk about it. This person might make an effort to avoid discussing himself, because he finds it embarrassing. Is such a person modest? No. Modesty is something that is partly internal, it is basically an attitude of ignorance which one has toward oneself which leads to modest patterns of behavior. The sort of person mentioned above is *simply* exhibiting a type of behavior, i.e., not talking about oneself, which can be correlated with modesty, but does not itself constitute modesty. If we regard behavior as a sufficient condition for modesty, then we will include cases where a person is insincere in exhibiting "modest" behavior; if we regard behavior as a necessary condition for modesty, then we will be leaving out cases where a person is modest, at least has modest dispositions, but who does not exhibit behavior that can be characterized as modest (though perhaps he does not act immodestly). The reason that most people believe modesty to be largely behavioral is that behavior is good evidence for telling whether a person is modest

[4] Aristotle does *not* consider this trait to be a virtue. When he is discussing the virtue of pride, he contrasts the proud man with the man who is unduly humble. This man is "[t]he man who thinks himself worthy of less than he is really worthy of . . ." (1123b9). Aristotle is bound to this view, however, because of his doctrine of the mean. The man who is unduly humble falls short of an accurate appraisal of himself, and thus fails to conform to the mean. Aristotle adopts condition (A), and rejects such traits as modesty, but this rejection does not sit well with the ordinary conception of virtue.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

or not. My point is simply that (modest) behavior should not be *equated* with modesty.

Even so, a response to this analysis will be that a person can be modest without talking about himself at all, where the person in question is not merely exhibiting certain behavior that constitutes the modesty. Rather, the behavior springs from an inclination to avoid discussing himself, even though he has an accurate assessment of his accomplishments and character. Is this not modesty? In my view, it is another form of false modesty, though it is not as ostentatious. Consider the analogy with lying once again. The sort of "modesty" described above would be like omissive lying. It is false modesty "by omission" rather than commission.

Nevertheless, this analysis could also be criticized for being too broad; that is, there might be cases which fulfill my conditions for modesty, but which we would not want to call examples of modesty. Consider, for example, a person who is involved in a very important project, but does not realize that his role is crucial, that he is doing a very good job, etc. But, if someone were to sit him down and *tell* him just how important he is, he would develop an inflated view of himself, because he is prone to conceit. Nevertheless, he does not realize how important he is. Most of us would not call this a case of modesty, because the person in question is disposed to conceit, and the fact that he does not realize how good he is does not erase that fact. What I want to say about modesty is that it is not enough to be ignorant of self-worth, one also has to be disposed to it. A truly modest person would simply not believe that he was that good. In a way, modesty can be characterized as a dogmatic disposition to underestimation of self-worth.

II.

In the first part of this paper, I discussed my views on modesty as a disposition to underestimate self-worth. This is how I want to characterize all sincere cases of modesty, but, as I pointed out at the beginning of the paper, I do not need this strong claim in order to be able to make my primary point. All I need is the extremely plausible claim that some cases of modesty involve underestimation of self-worth. Now I want to investigate what sort of virtue modesty is.

Modesty could be considered to be what Michael Slote[5] has chosen

[5] *Goods and Virtues* (New York: Oxford, 1983), pp. 61ff. Slote discusses humility as a virtue, and humility is very akin to modesty. The humble person has a tendency to avoid vanity, etc. But Slote discusses humility as though it were the same as modesty; he talks of humility as a virtue being dependent upon the humble person having other desirable traits, but, for example, choosing not to discuss them, or to take credit for them. But I believe that there is a distinction to be made here. A humble person, unlike a modest person, can paint an accurate, though perhaps unflattering picture of himself. The modest person needs to underestimate.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

to call a "dependent" virtue. A dependent virtue is a virtue only insofar as it is accompanied by some other virtue—the modest person has to be modest about *something*. Usually this something will involve some other desirable trait, like being hardworking or intelligent. This is important in evaluation. When we say "Jones is modest," we usually mean that Jones is modest with respect to some other good. Jones *may* be modest about everything, of course, but a disposition toward modest behavior does not necessitate this. Jones could be modest about his work, but not modest about his children, for example. This holds true for other virtues, I should think, since a parent could be cowardly about protecting himself, but display immense courage when his children are in danger. This merely illustrates that a person's virtues and vices often take objects, and those objects are many times other goods.

Modesty is a dependent ethical virtue, with an ethical value consisting in an underestimation of the self in some respect, or with respect to some desirable trait. But how does this underestimation have ethical value? To answer this question, I want to make use of a distinction between ethical considerations in evaluation, and non-ethical considerations—a distinction which has been discussed and elaborated by Bernard Williams.[6]

For Williams, the boundary surrounding and delimiting ethical considerations from other considerations is blurry. Notions such as "obligation" and "duty" are within the domain of ethical considerations, and virtues are also taken to be included in this domain. On Williams's account, when a person performs a virtuous act he rarely does it *only because* it is virtuous. Modesty fits this description, since the modest person cannot even know that he or she is being modest. Williams himself writes: "it is a notorious truth that a modest person does not act under the title of modesty" (*ibid.*, p. 10). Williams would therefore agree with the oddity of (1). This observation is true of other virtues as well—generosity, for instance. The generous person does not do the generous thing *because* he realizes it is generous; rather, he does the generous thing out of consideration for others. This observation should reinforce my account of modesty, but with a slight twist. With modesty, not only does the agent act without knowing he is modest, the agent does not even realize that he is underestimating himself. That is, the person cannot apply to himself any of the usual descriptions that others can apply to him. Others can say of him that he is underestimating his worth, and for this reason he is modest, but he cannot do this. The generous person does not act because of the generosity of the act, but presumably he can still

[6] *Ethics and the Limits of Philosophy* (Cambridge: Harvard, 1985), pp. 7 ff.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

describe the action accurately, and perhaps even realize that the act is generous while he is doing it. Thus, this analysis of modesty, and other virtues of ignorance, entails that an asymmetry exists between the self-ascription of the virtue, and the other-ascription of it. I can ascribe the virtue to another, but I cannot coherently and sincerely ascribe it to myself.[7]

A virtue is a disposition to act in a certain way, and usually not to perform an act merely because it is virtuous.[8] This observation alone should rule out false modesty as a virtue. This is because a person who understates self-worth is doing so because he believes it to be modest behavior. False modesty may be polite, or expedient, but it is not a moral virtue because it is too self-conscious an act. At the very best, it may be considered an aesthetic virtue in light of the fact that modest behavior is considered polite—but the understatement is not sincere underestimation. The point here is that false modesty can be used in an evaluation of an agent's character, but it is really evaluation of a nonethical sort, unless the person exhibiting it has unscrupulous motives for displaying modest behavior, or perhaps wholly altruistic ones.[9] This is because the falsely modest person exhibits the behavior he does because it conforms to expectations regarding modesty. There is something "put on" about false modesty which gives it the flavor of insincerity. There may be several explanations for why it has been considered a moral virtue. The explanation that I shall consider is derived from Aristotle:

> . . . the man who loves truth, and is truthful where nothing is at stake, will still more be truthful where something is at stake; he will avoid falsehood as something base, seeing that he avoided it even for its own sake; and such a man is worthy of praise. *He inclines rather to understate the truth; for this seems in better taste because exaggerations are wearisome* (emphasis added).[10]

Of course, Aristotle is talking here about understating the truth to avoid being considered a braggart. The element of Aristotle's view

[7] In Margaret Gilbert's "Vice and Self-knowledge," this JOURNAL, LXVIII, 15 (August 5, 1971): 443–453, this same sort of asymmetry is discussed with respect to evil.

[8] As Williams notes, there are exceptions to this observation, e.g., "just." A person can, and does, act according to justice by making decisions because they are just.

[9] But in these cases the person is not being blamed or praised for being falsely modest. Rather, the person is being blamed for being unscrupulous, or praised for being altruistic.

[10] *Nichomachean Ethics,* W. D. Ross, trans., in *The Basic Works of Aristotle,* Richard McKeon, ed. (New York: Random House, 1941), 1127b4–8. Ross translates the extreme of boastfulness as "mock-modesty," for Aristotle. This would indicate that, when Aristotle speaks of understating the truth (with respect to one's accomplishments), he also has in mind false modesty. See 1127b23–26.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

which I am considering here is this—that modesty (false or sincere) is considered a virtue, because it is negatively correlated with some behavior that is a vice. Exaggerating self-worth, for example, would be such a vice. In order to escape this vice, one might strive for modesty. And modesty is considered a virtue, because a person who is modest, even insincerely or falsely modest, does not exhibit the vice of exaggeration.

This account, however, is mistaken, because it conflates false modesty with sincere modesty. One virtue of my account is that it does treat the two as separate, since the sincere cases of modesty involve ignorance. Since modesty necessarily involves ignorance, it is also necessarily involuntary in nature.[11]

In understanding virtue, a distinction must be made between the cause of a particular trait being a virtue, and the object of the virtue (what the person has in mind in exercising the virtue). The modest agent is modest because he underestimates himself, and this leads to some good that is valued by those with whom he interacts (e.g., an easing of tensions, lack of jealousies, etc.). But the agent only intends to give an accurate assessment of himself, insofar as he is assessing himself at all.

III.

The above analysis illustrates the sort of virtue that I am interested in discussing here, and I suspect that the class of these character traits is really quite large. Another example is a virtue that I am going to call *blind charity*. This sort of charity is charity in thought rather than charity in deed. A person who is in blind charity with others is a person who sees the good in them, but does not see the bad. Blind charity differs from charity in that it is usually the case that, when one is merely charitable toward another, one favors that person in some respect, *in spite of* perceived defects or lack of desert. For example, in employing the principle of charity, one interprets a person's views in the best possible way, even though one perceives certain possible defects. Blind charity is a disposition not to see the defects, and to focus on the virtues of persons. A good example of a person exemplifying this virtue is found in the character of Jane

[11] Another possible way of dealing with virtues of ignorance is hinted at in Aristotle's text, when he writes that the truthful person tends to understate the truth because it is more tasteful than exaggeration. This would mean that the truthful person tends to be modest, because modesty is more pleasing than boastfulness. Perhaps it is possible to view the virtues of ignorance as aesthetic virtues rather than moral virtues. Thus, the analysis would be that, though (A) is necessarily true with respect to moral virtue, this is not the case for aesthetic virtue. Although I think that this is a very interesting line to pursue, it would entail developing an account of aesthetic virtues, and how they differ from moral ones—and this is an issue too large for the scope of this paper.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

 Digitized by Arya Samaj Foundation Chennai and eGangotri

Bennet in Jane Austen's novel *Pride and Prejudice*.[12] Jane is the sister of Elizabeth Bennet, the heroine of the novel. In the following bit of dialogue, Elizabeth and Jane are discussing Jane's possession of blind charity—Elizabeth finds it quite remarkable:

> 'Oh! you are a great deal too apt, you know, to like people in general. You never see a fault in anybody. All the world are good and agreeable in your eyes. I never heard you speak ill of a human being in my life'.
>
> 'I would wish not to be hasty in censuring anyone; but I always speak what I think'.
>
> 'I know you do; and it is *that* which makes the wonder. With *your* good sense, to be so honestly blind to the follies and nonsense of others! Affectation of candour is common enough; one meets it everywhere. But to be candid without ostentation or design—to take the good of everybody's character and make it still better, and say nothing of the bad—belongs to you alone . . .' (*ibid.*, pp. 14/5).

What can be said of such a character? Elizabeth, in this passage, seems to be mildly exasperated with Jane, but it is clear from the rest of the novel that she views Jane as one of the few truly good people in the world. And it is in large part due to Jane's sweetness, and her disposition to be blindly charitable.[13] This sort of charity is a charity which cannot be reflective—one cannot decide to be charitable in this way and grudgingly do it. It is a virtue much like modesty, because it requires that one be unknowing about something. Like modesty also, blind charity can be selectively directed, though Jane Bennet happens to be blindly charitable in general. Blind charity is therefore a virtue of ignorance, as is sincere modesty. The ignorance of blind charity is an ignorance about what is bad in other people, whereas modesty involves ignorance about the full extent of one's own goodness. These virtues of ignorance are involuntary in the sense that in a particular situation one does not choose, for example, to be charitable in the above sense, because one cannot know what to be charitable about. It may be that to cultivate these virtues one would need to practice self-deception. One might decide to cultivate a modest character, by cultivating certain beliefs about oneself. But, since these beliefs will have to be ones that the agent takes to be false, this process of cultivation is a process of self-deception. One might hold the position that modesty is a moral virtue where it arises spontaneously; it is only when it is cultivated that it cannot be considered

[12] New York: New American Library, 1961.

[13] Another character possessing total charity is the character of Melanie Wilkes in Margaret Mitchell's novel *Gone With The Wind*. Melanie is Scarlet O'Hara's rival for the affection of Ashley Wilkes. Interestingly, I have found few examples of male characters possessing this virtue, although Jane Bennet's beau in *Pride and Prejudice*, Mr. Bingley, is a close case.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

a moral virtue. This is a decidedly non-Kantian position to take. Here we would have a case where a person who tries to act virtuously necessarily fails at it. The person naturally disposed to be modest or blindly charitable succeeds.

Yet another trait that is a virtue of ignorance is a sort of refusal to hold a grudge. Imagine a situation in which Jones has harmed Smith —perhaps he broke a promise, or betrayed a trust. But Jones feels remorse, and genuinely desires that Smith forgive him, so that they can resume their friendship. And Smith wants to forgive him, so he tries to forget about the harmful incident. It may be the case that the forgetfulness is crucial to the sort of forgiveness to which Jones aspires. The old sayings, "forgive and forget" and "let's just forget about it," are not without significance. The person who feels that another has forgotten the harm is far more likely to be comfortable around that person, and to feel really forgiven. This is because one knows that the harm will not be conjured up to strain the relationship. Not just any forgetfulness is a virtue, however. It has to be conjoined with a plan of forgiveness, and an unwillingness to hold a grudge, even if the forgetfulness itself is not voluntary.

There are three general explanations for why we value the virtues of ignorance. The first is that there is some psychological state which we value intrinsically, and which necessarily involves ignorance. For example, in the case of modesty, the psychological state could be a reluctance to examine oneself thoroughly in terms of accomplishments—this reluctance to dwell on one's good qualities will lead to the ignorance associated with modesty. Thus, it is not the ignorance itself that is valued; rather, it is the psychological state that is accompanied by the ignorance.

The second explanation is instrumental. We value the virtues of ignorance because of the good that they generate. In the case of modesty, this will involve an easing of jealousy, envy, and tension between people. In surveying a sincerely modest person, harmful feelings like these are alleviated. This necessarily involves ignorance because, if we suspected that the person, in fact, was not ignorant, and thus insincere in his self–evaluation, his behavior would have the opposite effect. One might, for example, feel patronized. Here we value the effects, and the psychology does not seem so important.

The third explanation is an overlap of the first two. We value the virtues of ignorance because of the psychological states that underpin them, but we value these psychological states for instrumental reasons. This is the explanation which I consider to be the most plausible. Reluctance to take in one's own accomplishments fully, to avoid adding up one's worth, which leads to the ignorance of self-worth, could be valued because recognition of this trait in others

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

leads to an alleviation of the more destructive competitive emotions like jealousy and envy. Modesty oils the wheels, so to speak. It eases interaction between people, without being the product of convention. This is because modesty implicitly involves comparison. When the modest person makes the mistake, he is comparing his performance to the performance of others—and underestimating it. This is not his conscious program, in the sense that he does not think of himself as *comparing*. Nevertheless, the psychological mechanism that underpins the making of this sort of mistake is valued because the effects are valued. The modest person fails to elicit the sort of jealousy that a nonmodest person would engender. The modest person is, therefore, less troublesome.[14]

The person who wants to retain a condition like (A) as necessary for an account of moral virtue is now faced with the necessity of jettisoning a whole class of character traits which seem to be virtues. It is counterintuitive to suggest that traits like modesty and blind charity are not virtues. Indeed, I have made suggestions as to why we do find these traits valuable. A Nietzschean account of the virtues would have no problem in discounting these traits as virtues, but at the same time it would discount most of our traditional moral values. The virtues of ignorance do demonstrate that at least some element of a traditional account of the virtues is mistaken: either these traits are virtues, in which case (A) is false, or they are not, in which case they should not be cited in moral evaluation.

JULIA DRIVER

Virginia Polytechnic Institute and State University

[14] Thomas Nagel has suggested to me that an analogy might exist between this sort of an account of the virtues of ignorance and what Mohan Matthen calls "normal misperception," this JOURNAL, LXXXV, 1 (January 1988): 5–27. Normal misperception is due to our use of imperfect indicators, which give us a perceptual flexibility in discriminating a variety of colored surfaces under a variety of illuminations. Rare misperceptions are the price that we pay for this ability, since the indicators are the best available, though imperfect. An example that he discusses is the following: "It is possible to create a surface on which the color changes gradually, in just the way that illumination normally changes. In uniform illumination, such a surface would be perceived as uniformly colored and variably illuminated" (p. 12).

Virtues of ignorance are similar. It is not the ignorance itself which is valued. Rather, it is the underlying state that is necessarily connected to the ignorance in a certain type of context.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

BOOK REVIEWS

Vaulting Ambition: Sociobiology and the Quest for Human Nature. PHILIP KITCHER. Cambridge: MIT Press, 1985. xi + 456 p. Cloth $25.00, paper $12.50.*

The title of this book stems from some lines spoken by Macbeth, musing on regicide:

> . . . I have no spur
> To prick the sides of my intent, but only
> Vaulting ambition, which o'erleaps itself
> and falls on the other.

On one reading, Philip Kitcher plays a role like that of Macbeth; he aims to slay an unwitting victim, namely, "pop sociobiology." He would have us believe, however, that his victim, unlike Duncan, is utterly unworthy of a throne. Pop sociobiology, though naive to the extreme, is not innocent; it is poor science, fraught with unjustified and socially consequential claims. Furthermore, it is ill-equipped to defend itself in a fair fight, for it is easily vanquished by sound and well-sharpened methodological and philosophical weapons.

There is a certain tension between rhetoric and argument in Kitcher's account of his intended victim; if pop sociobiology is truly as unworthy as the former suggests, this is no regicide and one should not need some 440 pages of often dense argumentation to carry out the foul deed. Why all these heavy cannons? Implicitly, Kitcher offers at least two responses: first, the victim is noxious; protected by mimicking sound ethological work, it is a diseased rat which has been abroad in the land for a considerable time, spreading pestilence and plague. Clearly, then, the job is important even if the victim is not worthy.[1] Second, there are a number of useful animals in the intellectual woods bearing a superficial resemblance to our rat. These deserve protection and admiration. Care is required, therefore, to discriminate properly between the rat and its relatives.

* I am grateful to Marjorie Grene, Joseph Pitt, and Arthur Squires for helpful comments on a draft of this review, and to Philip Kitcher for suggesting some improvements in the final version.

[1] In this respect, the present volume is reminiscent of Kitcher's earlier anti-creationist book, *Abusing Science* (Cambridge: MIT, 1982). *Vaulting Ambition,* unlike its predecessor, develops a number of serious internal criticisms of current evolutionary theory, exploring, for example, the problem of excessive adaptationism.

0022-362X/89/8607/385–391 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

I shall not attempt to sustain these metaphors. But they are useful foils in appraising this ambitious book, which offers both a polemic against the excesses of pop sociobiology and an attempt to show what is sensible and worthwhile in the less ostentatious sociobiological work on the evolution of animal behavior. In the course of these pursuits, Kitcher deals with the foundations of evolutionary biology, difficulties of certain familiar reductionist approaches to the study of humans, and crucial issues in discriminating good science from bad. In his concluding chapter, he seeks, further, to set limits on the use of current biology to yield pronouncements regarding moral value, right and wrong, and human freedom. It is not possible to deal with all of these topics in this review.

Let us turn to a survey of the book. The introduction addresses the political implications of popularizations of sociobiology, starting from Kitcher's keenly personal awareness of the detrimental effects of the "eleven-plus"s (the examinations that, until recently, determined the educational tracking of children in England) and of the use of scientifically inadequate evidence to justify them. Kitcher argues that much of the dispute about sociobiology concerns the adequacy of certain bodies of evidence and maintains that "what counts as sufficient evidence is not independent of the political consequences" (9). The violence of Kitcher's rhetoric throughout[2] confirms what is clear from the beginning: he holds that pop sociobiology does not meet the minimum standards of evidence required of serious science, let alone the higher standards appropriate to politically consequential views propounded with the authority of science. He offers powerful arguments in support of this stance; I shall return to consider precisely how far those arguments reach.

Chapter 1 identifies pop sociobiology as a cluster of recent work "appealing to recent ideas about the evolution of animal behavior in order to advance grand claims about human nature and human social institutions" (14/5). The principal targets of Kitcher's polemic are E. O. Wilson, his co-author C. Lumsden, Richard Alexander, and their allies. Chapters 2–3 provide an account of the main features of evolutionary theory and of some recent developments in evolutionary biology of particular relevance for Kitcher's purpose. Chapter 4 sets forth a serious account of the content of sociobiology. "*Broad* sociobiology" (114/5) fits Wilson's definition of the field as "the

[2] To cite but one example, the comparison of Richard Alexander's argumentative strategy with that of the creationist Henry Morris (298) implicitly condemns Alexander's work as unscientific in a far stronger vein than seems justified. This issue will recur at the end of this review.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

systematic study of the biological basis of all social behavior."[3] It includes studies of the functions of behaviors; the mechanisms that produce and the stimuli that trigger them; the developmental trajectories by which organisms come to acquire behaviors, behavioral repertoires, and behavioral dispositions; and the evolutionary histories of all of these. Kitcher subsequently argues that all attempts to employ current developments in broad sociobiology in order to ground a doctrine of human nature, to provide a foundation for ethics, or to develop an account of biological constraints on social planning or amelioration are methodologically unsatisfactory and socially and philosophically unwarranted—indeed, that the available evidence shows all attempts to base the study of *human* social behavior on the doctrines of pop sociobiology are wholly illegitimate. In contrast, "*narrow* sociobiology" (115 ff.) focuses mainly on evolutionary questions, primarily about nonhumans, often with particular concern for the adaptive value and evolutionary history of behaviors; on the whole, its work is better grounded than that of broad sociobiology, but it has few direct consequences for the issues that most concern pop sociobiologists. Furthermore, narrow sociobiology is not likely to yield a Wilsonian "new synthesis" because "there is no general theory of the evolution of animal social behavior that is distinct from the general theory of evolution" (118).

Chapters 5–8 are the methodological core of the book; they compare work on nonhumans with that on humans, examine the adaptationist strategies built into much sociobiological argumentation, provide an account of the theoretical techniques employed in the current literature, develop particular methodological issues, and apply all of these considerations to some of the principal cases that Wilsonian sociobiologists employ in support of their theories and points of view. These include, for example, sociobiological treatments of xenophobia and homosexuality, and of the prevalence of certain sex roles and of incest taboos in nearly all cultures. Of particular interest is Kitcher's repeated demonstration here and in chapter 10 of specific errors in pop sociobiological attempts to fill in what he calls "Wilson's ladder," according to which such prevalent behaviors are (1) adaptive, (2) produced and maintained by natural selection, and (3) acting on genetic differences among people and populations, so that (4) it will be difficult to modify those behaviors by means of available environmental interventions. The arguments over cases are

[3] Quoted (at 113) from E. O. Wilson, *Sociobiology: The New Synthesis* (Cambridge: Harvard, 1975), p. 2.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

not always easy, but they are nearly always on target and require serious attention by the advocates of sociobiology.

Chapters 9–10 focus on two leading pop sociobiological books—one by Alexander[4] and the other by Lumsden and Wilson.[5] The final chapter addresses the bearing of (pop) sociobiology on human altruism, human freedom, and the foundations of morality. Very roughly, biology can supply factual premises, but not moral principles or values. (Kitcher's argument to this effect seems to me definitive and incontrovertible.) Humean compatiblism regarding freedom can be defended against all available biological counterarguments.

Among Kitcher's themes, I find the treatment of good versus bad science the most troublesome. In a thoroughly persuasive, though difficult, chapter he demonstrates Lumsden and Wilson's failure to provide satisfactory arguments for the major conclusions of their book. Kitcher maintains that their complex mathematical apparatus could (in principle) serve as a vehicle for the evaluation of the effects of the interactions between genetic and cultural systems of inheritance or transmission of social traits, behaviors, or dispositions. But, as he shows, the gap between principle and practice, between available knowledge and that required to determine the parameters of Lumsden and Wilson's equations, is enormous. For one thing, the fancy mathematics sometimes hides the fact that the very same results can be derived from far less controversial theories, with much less reliance on onerous and poorly supported assumptions. (Kitcher demonstrates this well for such conclusions as "Pure tabula rasa is an unlikely state," i.e., that there will be genetic biases affecting dispositions to social behavior (L & W, 290) and "Culture slows the rate of genetic evolution" (L & W, 294).) Other conclusions, though legitimately derivable from Lumsden and Wilson's apparatus when certain parameters are assigned, depend on arbitrary assumptions which, when dematheticized, are utterly implausible. Simplifying Kitcher's translation of one such, Lumsden and Wilson show that groups of people who imitate the behavior of those around them, thereby adjusting their choices about a culturally significant behavior affecting fitness, will evolve a genetic bias in favor of fitness-promoting behaviors faster than groups without such imitative skill. The trouble with the "proof" of this claim is that it assigns the relevant parameters so as to levy, in effect, the *entire* cost of building a brain capable of imitation against the choice between *two* opposed behav-

[4] *Darwinism and Human Affairs* (Seattle: Washington UP, 1979).
[5] *Genes, Mind, and Culture* (Cambridge: Harvard, 1983) (hereafter L & W).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

iors ("culturgens" in Lumsden and Wilson's terms; cf. Kitcher, 379–386). To put it mildly, it is not obvious that the idealized assumptions involved here can be replaced by more plausible ones without doing irreparable damage to the argument.

Kitcher is utterly persuasive about such cases. He has shown convincingly that a vast panoply of technical arguments do not lend solid support to the nontechnical conclusions of pop sociobiologists or to the Wilsonian "picture" of human evolution and the biological constraints it imposes on social change. In this important sense, pop sociobiology is bad science. Indeed, this is the heart of Kitcher's complaint about pop sociobiology in general; the specific evolutionary scenarios it has offered to date are ill–supported and, even if correct, do not justify the suggestions of genetic constraint on social behavior which its advocates are prone to draw.

Kitcher draws a questionable conclusion from this just complaint. He challenges the sociobiological program, if it is able, to replace its impoverished arguments about human sociobiology with better ones that will withstand critical scrutiny, but his rhetoric asks us to dismiss altogether pop sociobiological attempts to deal with humans. Thus, in the present instance, Kitcher concludes that mathematics is used by Lumsden and Wilson "to disguise the poverty of thought" (393), that the novel results are merely a matter of "number juggling," "artifice," and "trickery" (392). Fair enough, if this is meant to call attention to the gap between the technical arguments and the grand claims about human nature. Kitcher seems, however, unfairly to forget what he himself allows (e.g., at 337/8)—that such equations are, in principle, the proper tool for exploring gene-culture interactions. Kitcher's own exposition of the sensitivity of social evolution to strictly environmental differences exploits the very models developed by Lumsden and Wilson but misused in support of their larger views.

The consequences of Kitcher's criticisms for the sociobiological research program (as opposed to their specific claims about human nature and behaviors) are, in general, weaker than his rhetoric implies. On his own account, theories should be understood in terms of the questions they address, the strategies they offer for answering those questions, and the general ideas they offer about the domain with which they are concerned (54). The questions at stake here are surely legitimate, although it is not obvious whether any theory will prove capable of handling them well. Lumsden and Wilson's strategy, according to Kitcher himself, is appropriate in principle for dealing with human sociobiology; the fact that they do not execute

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

that strategy well means that we must reject their arguments, but leaves the ultimate evaluation of that strategy open.

It is, of course, extremely important to recognize that Lumsden, Wilson, Alexander, and their allies have failed to justify their picture of the consequences of our evolutionary history. This failure, driven home by Kitcher in rich detail, provides the primary reason that social policy should not be based on their views. Nonetheless, it would be a mistake to dismiss pop sociobiology altogether as "bad science," resting merely on artifice and impoverished thinking, methodologically more-or-less on a par with creationism. We simply are not yet in a position to know what picture will emerge from continuing efforts to apply evolutionary theory to humans. And, as Kitcher's official argument recognizes (and as he himself will admit in his pluralist moments), the strategy and techniques on which his opponents draw are legitimate even if the application is disastrously flawed.

Indeed, I would urge strongly the value of a pluralistic science here. Workaday science has important room for "stories,"[6] that is, for partial pictures, often poorly justified and articulated. Many important research programs rest on such stories, which serve to motivate research and drive the processes of "paradigm articulation" and theory testing. These processes are, properly, comparative. For this reason, "stories" that are implausible but powerful often play an important role in fundamental research; they motivate some to attempt to show that they are true and valuable and others to refute them. In this way they serve as important incentives for the improvement and testing of opposing theories.

To this extent, at least, the best way to avoid the misuse of science in support of "eleven-plus"s or other social reforms, however intended, is not to ban "stories" from science, but to insist on a very high standard of support before they are taken seriously in debates over social policy. We must refine our distinctions among "stories" —some are mere exploratory speculations, others have some degree of explanatory value over a limited range, and a very few are well enough supported to serve as a basis for social policy or for socially consequential actions. Kitcher's powerful arguments show that the general picture of human nature developed by pop sociobiologists should be classed as mere speculation—but he has not yet justified

[6] I have taken this use of the term 'story', with permission, from helpful discussions with Arthur Squires, whose draft manuscript, *The Left Hand of Love*, bears usefully on these issues.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

barring the path of inquiry by dismissing their "stories" out of hand. For this reason, I suggest, we are not yet ready to do what Kitcher's rhetoric suggests he has done, namely, to "put an end to the controversies that have beset sociobiology in the past decade" (ix).

In discussing the risk of premature dismissal, I have focused on an overextension of Kitcher's argument. This should not obscure the fact that his book has many virtues inadequately discussed here. Above all, he offers a rich banquet of insights into the conceptual geography of sociobiology (and evolutionary theory generally) and an extraordinary wealth of critical arguments with which the proponents of human sociobiology will have to cope if their program is to survive. Some of his arguments, like those revealing the limitations of falsifiability as a mark of the value of theories (58 ff.), will be of special value to biologists. Others, including the general account of the structure of evolutionary theory and the place of "Darwinian histories" in that theory, are of importance for philosophers of biology. An especially nice argument of general interest concerns the blindness of Lumsden and Wilson's reductionist program to the difference between *social institutions* (like those promulgating incest taboos in many societies) and the sum of individual dispositions to behave in certain ways (e.g., to avoid incest, perhaps with distaste, or to berate or punish others for incestuous acts) (344–350). The concluding chapter can be read independently of the rest of the book; it is a useful defense of the autonomy of moral principles and human freedom (as analyzed along Humean lines) vis-à-vis the findings of science. These virtues make the book highly worthwhile. It is, therefore, a pity that Kitcher's exaggerated rhetoric and the necessarily technical character of a considerable part of his argument are likely to cause many of his readers to miss the beauty and effectiveness with which he wields an argumentative scalpel when he is at his best.

RICHARD M. BURIAN
Virginia Polytechnic Institute and State University

Johnsonian Prize Announcement

The Editors of the JOURNAL OF PHILOSOPHY regret to announce that, for reasons of scheduling, the Johnsonian Prize which we sponsor with Hackett Publishing Company will be cancelled for 1989. Any manuscripts that authors were planning to submit this year will be considered for the 1990 Prize, the deadline for which remains January 15, 1990. We apologize for the inconvenience this cancellation may cause any authors and sponsoring departments, and we look forward to reviewing your manuscripts for the 1990 Prize, the decision for which will be announced in October 1990.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

 Digitized by Arya Samaj Foundation Chennai and eGangotri

NOTES AND NEWS

Harvard University announces the Andrew W. Mellon Faculty Fellowships in the Humanities. The fellowships are available for nontenured, experienced junior scholars who have completed, at the time of appointment, at least two years of postdoctoral teaching, as college or university faculty in the humanities. Further information may be obtained by writing Richard Hunt, Harvard Univ. Mellon Faculty Fellowships, Lamont Library 202, Cambridge, MA 02138.

The Center for Philosophy of Religion at the University of Notre Dame is pleased to announce the Senior and Junior Fellowships for the 1990/1 academic year. The aim of the fellowships is to enable holders to pursue research and writing projects in the philosophy of religion, particularly Christian philosophy. The Distinguished Scholar Fellowships are intended to provide time for reflection and writing to those whose work is in the forefront of current research. The Postdoctoral Fellowships are offered to those whose tenure at the Center would allow them to make progress in the Center's areas of interests. Fellows will be asked to teach one class per semester, and will have access to university facilities. The deadline for Distinguished Scholar application is December 1, and for Postdoctoral Fellowships is April 1, 1990. Further information may be obtained by writing Alvin Plantinga, Center for Philosophy of Religion, 330 Decio Hall, Box 1068, Notre Dame, IN 46556.

The Center for the Critical Analysis of Contemporary Culture at Rutgers University is pleased to announce an international conference on "Realism and Representation," to be held on November 10–12. The conference will consider the possible relations between science, the philosophy and history of science, and literary theory and criticism. Speakers will include Svetlana Alpers, Jonathan Arac, Gillian Beer, James Clifford, Bas van Fraassen, Katherine Hayles, Evelyn Fox Keller, Ludmilla Joranova, Paisley Livingston, J. Hillis Miller, Richard Miller, Harriet Ritvo, Richard Rorty, Simon Schaffer, and Robert Scholes. Further information may be obtained by writing Realism and Representation, CCACC, 180 College Ave., New Brunswick, NJ 08903.

The Pacific Division of the Society for Women in Philosophy is pleased to announce the Second Annual Merrill Hintikka Prize Essay Competition. The prize will be awarded to one faculty member and one student for the best essay in feminist philosophy. The papers must be previously unpublished and will be read at the APA Pacific Division Meetings in Spring 1990. Papers, which must be received by September 1, may be directed to Rita Manning, Phil. Dept., San Jose State Univ., 1 Washington Sq., San Jose, CA 95192.

The Department of Philosophy of the University of Tennessee/Knoxville is pleased to announce the formation of the Center for Applied and Professional Ethics (CAPE), under the direction of Glenn Graber. CAPE develops programs of education, conducts conferences, and consults on matters of professional and applied ethics. Information may be obtained by writing Glen Graber, Phil. Dept., Univ. of Tennessee, Knoxville, TN 37996.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVI, NO. 8, AUGUST 1989

REALISM AND THE HISTORICITY OF KNOWLEDGE

THE problem described in the title is not new. In the West it arose with the Presocratics; it was formulated by Plato and Aristotle, dismantled by the rise of modern science, and reappeared with quantum mechanics and the increasing strength of historical (as opposed to theoretical) accounts of knowledge. Stated briefly, the problem consists in the following question: How can information that is the result of idiosyncratic historical changes be about history-independent facts and laws? To examine the problem, I shall replace it by two assumptions and the difficulties that arise from their common use.

I. THE ASSUMPTIONS

The first assumption is that the theories, facts, and procedures that constitute the (scientific) knowledge of a particular time are the results of specific and highly idiosyncratic historical developments.

Many events support this assumption. The Greeks had the mathematics and the intelligence to develop the theoretical views that arose in the sixteenth and seventeenth centuries, yet they failed to do so. "Chinese Civilization," writes J. Needham,[1] "had been much more effective than the European in finding out about Nature and using natural knowledge for the benefit of mankind for fourteen centuries or so before the scientific revolution," and yet this revolution occurred in "backward" Europe: deficient, not good, knowledge led to better knowledge. Babylonian astronomy concentrated on special

[1] *Science in Traditional China* (New York: Cambridge, 1981), pp. 3, 22ff. For details, cf. the sections of Needham et al., *Science and Civilisation in China* (New York: Cambridge, 1956ff), esp. vol. v, pt. VII. A brief but suggestive comparison of simultaneous developments in Greece and China is found in the articles by Jacques Gernet and Jean-Pierre Vernant in Vernant, *Myth and Society in Ancient Greece*, (New York: Norton, 1988), pp. 79ff.

0022-362X/89/8608/393–406

© 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

events, such as the first visibility of the moon after a new moon, and constructed algorithms for predicting them. Trajectories, a celestial sphere, and the considerations of spherical trigonometry played no role in these algorithms. Greek astronomy posited first physical (Anaximander), then geometrical trajectories, and built on them. Both methods were empirically adequate and capable of refinement (equants, excenters, epicycles in Greek astronomy, polygons instead of step functions or zigzag functions in the Babylonian alternative). Cultural factors, not empirical adequacy, determined the survival of the one and the disappearance of the other.[2]

Recent studies added impressive evidence to these general observations. Historians of science examining the microstructure of scientific research, especially of modern high-energy physics, found many points of contact between the establishment of a scientific result and the conclusion of a complicated political treaty. Even experimental "facts" turned out to depend on compromises between different groups with different experiences, different philosophies, different financial backing, and different bits of high theory to support their position. Numerous anecdotes confirm the historico-political nature of scientific practice.[3]

The second assumption is that what *has been found* in this idiosyncratic and culture-dependent way (and is therefore formulated and explained in idiosyncratic and culture-dependent terms) *exists* independently of the circumstances of its discovery. We can cut the way from the result without losing the result. I shall call this assumption the *separability assumption.*

The separability assumption, too, can be supported by a variety of reasons. Indeed, who would deny that there were atoms long before the scintillation screen and mass spectroscopy, that they obeyed the laws of quantum theory long before those laws were written down, and that they will continue to do so when the last human being has disappeared from earth? And is it not true that the discovery of America, which was the result of political machinations set in motion

[2] The idea that Greek astronomy was inherently superior reflects part of the evidence which has stable theories on the one side and a rapidly growing research programme on the other. But the growth here and the stability there did not come from empirical successes or obstacles; it came from different social conditions; cf. G. E. R. Lloyd, *The Revolution of Wisdom* (Berkeley: California U P, 1987), ch. 2.

[3] Details in Andrew Pickering, *Constructing Quarks* (Chicago: University Press, 1984); Peter Galison, *How Experiments End* (Chicago: University Press, 1987); and the September 1988 issue of *Isis*. Anecdotes are unpopular among logicians who see science propelled by the rational use of rational principles. They become important once scientific practice has been reintegrated with history.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

by false beliefs and erroneous estimates and which was misread by the great Columbus himself, did in no way affect the properties of the continent discovered? It does not matter that some philosophers have objected to sweeping judgments such as these and have urged us to let science speak for itself, for "science itself" is full of sweeping judgments such as those just mentioned.[4]

So far we have seen the two assumptions that underly the problem; now for the difficulties.

II. THE DIFFICULTIES

The separability assumption is not only part of science but also of nonscientific traditions. According to Herodotus and Greek commonsense (sixth and fifth centuries B.C.), Homer and Hesiod did not create the gods, they merely enumerated them and described their properties. The gods had existed before and they were supposed to live on, independently of human wishes and mistakes. The Greeks, too, thought they could cut the way from the result without losing the result. Does it follow that our world contains particles and fields side by side with demons and gods?

It does not, the defenders of science reply, because gods do not fit into a scientific worldview. But, if the entities postulated by a scientific worldview can be assumed to exist independently of it, then why not anthropomorphic gods? True, few people now believe in such gods and those who do only rarely offer acceptable reasons; but the assumption was that existence and belief are different things and that a new Dark Age for science would not obliterate atoms. Why should the Homeric gods—whose Dark Age is now—be treated differently?

They must be treated differently, the champions of a scientific worldview reply, because the belief in gods did not just disappear; it was removed by argument. Entities postulated by such beliefs cannot

[4] Arthur Fine, in an interesting and highly informative book, *The Shaky Game* (Chicago: University Press, 1986), invites us "to let science stand on its own and to view it without the support of philosophical 'isms' " (9): "try to take science on its own terms, and try not to read things into science" (149). Having proposed a similar view in *Against Method* (London: New Left, 1975), I have much sympathy for Fine's position, but I cannot accept it as final. Science is not the only enterprise producing existence claims, and scientific arguments not the only "complex network of judgements" to "ground" such claims (153). Are we to become complacent relativists and accept as existing whatever people lay on us in a sufficiently complicated way ("*complex* network of judgements"), or should we not rather choose among networks and find reasons for our choice? At any rate, there are enough "objective" existence claims contained in molecular biology, the theory of evolution, cosmology, even in high energy physics (cf. Pickering, *op. cit.*, p. 404, on the "retrospective realism" of scientists) to bring the difficulties of the next section right into the center of science.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

be said to exist separately. They are illusions, or "projections"; they have no significance outside the projecting mechanism.

But the Greek gods were not "removed by argument." The opponents of popular beliefs about the gods never offered reasons that, using commonly held assumptions, showed the inadequacy of the beliefs. What we do have is a gradual social change leading to new concepts and new stories built from them.

To see this, consider two early objections against the gods of Homer. The first objection occurs in a well-known lampoon by Xenophanes,[5] the traveling philosopher. Ethiopians, he says, picture their gods as snubnosed and black; Thracians as blue-eyed and blond. And he adds:

> But if cattle, or lions, or horses had hands, just like humans;
> If they could paint with their hands, and draw and thus create pictures
> then the horses in drawing their gods would draw horses; and cattle
> would give us pictures and statues of cattle; and therefore
> each would picture the gods to resemble their own constitution (*ibid.*, fragment 15).

In contrast, the "true" god is described as follows:

> One god alone is the greatest, the greatest of gods and of men
> not resembling the mortals, neither in shape nor in insight.
> Always without any movement he remains in a single location
> as it would be unseemly to walk now to this, now to that place.
> Totally vision; totally knowledge; totally hearing.
> But without effort, by insight alone, he moves all that is (*ibid.*, 23–26).

Can these lines be expected to convert a convinced "Ethiopian" or "Thracian"? Not a chance! For the obvious reply is: "You, Xenophanes, seem to dislike our gods—but you have not shown that they don't exist. What you did show is that they are tribal gods, that they look and act like us and that they differ from your own idea of god as a superintellectual. But why should such a monster be a measure of existence"?

Xenophanes's mockery can even be inverted, as we see from a remark by Timon of Phleios, a pupil of Pyrrho:

> Xenophanes, semi-pretentious, made mincemeat of Homer's deceptions,
> fashioned a god, far from human, equal in all his relations,
> lacking in pain and in motion and better at thinking than thought.[6]

[5] In *Fragmente der Vorsokratiker*, Diels-Kranz, eds. (Zürich: Weidmann, 1985), fragment 16.

[6] In Sextus Empiricus, *Outlines of Pyrrhonism*, I, 224.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The upshot is that Xenophanes's mockery works only if the entity he wants to introduce already impressed itself on the minds of his contemporaries; *it formalizes a historical process, it cannot bring it about.* (I suspect that almost all "arguments" that "advance thought" have this feature.)

A second early "proof" against polytheism which was recorded by the Aristotelian school[7] makes the situation even clearer. According to the proof,

god is either one or many
if many, then either equal or unequal
if equal, then like the members of a democracy
but gods are not members of a democracy, hence unequal
but if unequal, the lesser is not a god
hence
god is one.

The proof assumes that being divine means the same as having supreme power. This was not true of the Homeric gods. Again, the proof hits its aim only after the needed change in the notion of divinity has occurred: *history, not argument, undermined the gods.* But history cannot undermine anything, at least not according to the separability assumption. This assumption still forces us to admit the existence of the Homeric gods.[8]

It does not force us, scientific realists object, because a belief in anthropomorphic gods, though perhaps not removed by reason, never was a reasonable belief. Only entities postulated by reasonable beliefs can be separated from their history. I shall call this the *modified separability assumption.*

Now, making reasonability a measure of separable existence means assuming that things are adapted to criteria of existence and not the other way around. Scientific practice does not conform to this as-

[7] In *On Melissus, Xenophanes, Gorgias,* 977 a14ff.

[8] Marcello Pera objected that I used too narrow a notion of "argument" to make my point. I reply that by 'argument' I mean any story that can be told in a relatively short time, has the purpose of showing that the Homeric gods do not exist, and does this in an "intellectual" way, i.e., using propositions and not procedures such as terror, hypnosis, etc.

It should also be added that the ancient gods were not creatures of fantasy, but a living presence. In hardhearded Rome, they even participated in the political process. Cf. *Against Method,* rev. ed. (London: Verso, 1988), ch. 16 (Homeric gods); Robin Lane Fox, *Pagans and Christians* (New York: Norton, 1987), pt. 1, sect. 4 (title: "Seeing the Gods"—late Roman Empire); and Donald Strong, *Roman Art* (London: Penguin, 1982) (Roman Republic and Empire through the fourth Century). The triune God of Christianity and the saints profoundly influenced the arts, the philosophy, and the politics of the West and this not only in retrospect, but for those who worked out the details.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

sumption, and rightly so. Birds are said to exist because we can see them, catch them, and hold them in our hand. The procedure is useless in the case of alpha-particles, and the criteria used for identifying alpha-particles do not help us with distant galaxies or with the neutrino. Quarks, for a time, were a doubtful matter, partly because the experimental evidence was controversial, partly because new criteria were needed for entities allegedly incapable of existing in isolation ("confinement"). We can measure temperature with a thermometer; but this does not get us very far. The temperature in the center of the sun cannot be measured with any known instrument, and the temperature of events such as the first few seconds of the universe was not even defined before the arrival of the second law of thermodynamics. In all these cases, criteria were adapted to things and they changed and proliferated when new things entered the stage. To say that the Homeric gods do not exist because they cannot be found by experiment or because the effects of their actions cannot be reproduced violates this procedure. For if Aphrodite exists, and if she has the properties and idiosyncracies ascribed to her, then she certainly will not sit still for something as silly and demeaning as a test of reproducible effects (shy birds, people who are easily bored, and undercover agents behave in a similar way).

Let me dwell a little on this point. "A wise man," says Hume[9] (and many scientists agree), "proportions his belief to the evidence." Aristotle elaborates:

> Of all beings naturally composed some [the stars] are ungenerated and imperishable for the whole of eternity, but others are subject to coming-to-be and perishing. It has come about that in relation to the former . . . the studies we can make are less, because both the starting points of the inquiry and the things we long to know about present extremely few appearances to observation. We are better equipped to acquire knowledge about the perishable plants and animals because they grow beside us. Both studies have their attractions. Though we grasp only a little of the former, yet because the information is valuable we gain more pleasure than from everything around us, just as a small and random glimpse of those we love pleases us more than seeing many other things large and in detail. But the latter, because the information about them is better and more plentiful, *take the advantage in knowledge. . . .*[10]

But the epistemic power ascribed to areas of research does not conform to this principle. Artisans at all times possessed detailed information about the properties of materials and of their behavior under

[9] *An Enquiry Concerning the Understanding*, sect. x.
[10] Aristotle, *De Partibus Animalium*, I, 5.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the most varied of circumstances, whereas theories of matter from Democritus to Dalton were considerably less specific and their connections with the evidence much more tenuous.[11] Yet questions of reality and of suitable methods of discovery were often formulated in their terms, not in artisan terms. The information of artisans did not even count as knowledge. More recently, hydrodynamics and the theory of elasticity, though connected with experience in many ways, received a place far below abstract mechanics (Lagrange, Hamilton). Even comprehensive sciences, such as chemistry and biology, for a long time were assigned a secondary role in the hierarchy of the sciences. When the time scale of mid-nineteenth-century geology and biology exceeded the estimates of the solar age as determined by physicists (such as Helmholtz and Kelvin) and corresponding estimates of the cooling of the earth's surface, prevalence was given to the highly conjectural numbers of the latter.[12] We have to conclude that the authority of a subject, like its shape, is as much a result of idiosyncratic historical developments. True, the persistence of the atomists paid off: modern elementary particle physics, quantum chemistry, and molecular biology would not exist without it. But these achievements could not be foreseen, and an appeal to unknown and as yet inconceivable future effects can also be applied to the gods. Thus, the modified separability assumption cannot help us to get rid of gods either: having decided to separate history and existence, we must separate the existence of the gods even from the most "advanced" scientific argument.

Result: neither the separability assumption nor the modified separability assumption can make us accept atoms but deny gods. A realism that separates being and history and yet assumes that being can be grasped by history is forced to populate being with all the creatures that have been considered and are still being considered by scientists, prophets, and others. To avoid this abundance, some philosophers and scientists made the following move: Scientific entities

[11] Cyril Stanley Smith [*A Search for Structure* (Cambridge: MIT, 1981)] distinguishes between theories of matter such as the atomic theory and a knowledge of materials. He describes how the latter arose millenia before the former, was more detailed, and was frequently impeded by theoretical considerations. In an exhibition [items presented and explained in *From Art to Science* (Cambridge: MIT, 1980)] he showed the enormous amount of information contained in the products of early artisans. Norma Emerton [*The Scientific Reinterpretation of Form* (Ithaca: Cornell, 1984)] describes the battle between form theories (which were fairly close to the practice of the crafts) and atomism (which was not) and comments on the methods the atomists used to remain on top.

[12] J. D. Burchfield, *Lord Kelvin and the Age of the Earth* (New York: Hill & Wang, 1975).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(and, for that matter, all entities), they said, are projections and thus tied to the theory, the ideology, and the culture that postulates and projects them. The assertion that some things are independent of research, or history, belongs to special projecting mechanisms that "objectivize" their ontology; it makes no sense outside the historical stage that contains the mechanisms.[13] Abundance occurs in history, it does not occur in the world.

But not all projections are successful. The "scientific entities" mentioned above are not simply dreams; they are inventions that went through long periods of adaptation, correction, and modification, and then allowed scientists to produce previously unknown effects. Similarly, the gods of antiquity and the triune God of Christianity who replaced them were not just poetic visions. They also had effects. They influenced the lives of individuals, groups, and entire nations. Gods and atoms may have started as "projections"—but they received a response, which means they apparently succeeded in bridging the gulf that naive realists had erected between being and their own historical existence. Why has this response disappeared in the case of (the) gods? Why is it so powerful in the case of matter?

III. THE POWER OF SCIENCE

What I have said so far is simple and straightforward. Even a well-constructed argument cannot remove the overwhelming impression, however, that anthropomorphic gods were killed by the arrival, first, of philosophical and, then, of scientific rationalism. I shall therefore approach the matter again, and from a slightly different angle. Gods —and this now means the Homeric gods as well as the omnipotent creator-god of Christianity—are not only moral, but also physical powers. They cause thunderstorms, earthquakes, and floods; they break the laws of nature to produce miracles; they raise the sea and stop the sun in its course. But such events are now either denied or accounted for by physical causes, and remaining lacunae are being swiftly closed by research. Thus projecting the theoretical entities of science, we remove gods from their position of power and, as the more fundamental scientific entities obey time-independent laws, show that they never existed. Many religious people have accepted this argument and diluted their creed until it agreed with this philosophy.

Now, the fact that science dominates certain areas of knowledge does not by itself eliminate alternative ideas. Neurophysiology provides detailed models for mental processes; yet the mind-body prob-

[13] Fine has shown that this was also the way in which Einstein understood his own "realism" (*op. cit.*, ch. 6).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

lem is being kept alive, both by scientists and scientifically inclined philosophers. Some scientists even demand that we "put[ ] mind and consciousness in the driver's seat,"[14] i.e., that we return to them the power they had before the rise of a materialistic psychology. These scientists oppose the elimination and/or reduction of pre-scientific psychological ideas and entities. There is no reason why gods whose numinous aspects always resisted reduction should be treated differently.

Secondly, a reference to basic time-independent laws works only if the modern accounts of divinely caused events such as thunderstorms, earthquakes, vulcanic eruptions, etc., can be reduced to them. But there exist no acceptable reductions of the required kind. Special fields introduce special models whose derivability from basic physics is assumed, not shown. Already Descartes was aware of this situation. Commenting on the richness of the world, he professed himself unable to reduce the properties of special processes such as light to his own basic principles and used a variety of "hypotheses" instead.[15] Newton's[16] discussion of the properties of motion in resisting media abandons the deductive style of his planetary astronomy. It is "almost entirely original and much of it is false. New hypotheses start up at every block; concealed assumptions are employed freely and the stated assumptions sometimes are not used at all."[17] Modern researchers in the area explicitly reject reductive demands: "It is as ridiculous to deride continuum physics because it is not obtained from nuclear physics as it would be to reproach it with lack of foundation in the Bible."[18] The general theory of relativity for a long time was connected with the known planetary laws by conjecture, not by derivation (calculating the path of Mercury, one added the Schwarzschild solution to the results of prerelativistic perturbation theory without having shown from basic principles that both adequately described the situation in the planetary system). The connection between quantum mechanics and the classical level is quite obscure and has only recently been dealt with in a more satisfactory manner.[19] Meteorology, geology, psychology, large parts of

[14] R. Sperry, *Science and Moral Priority* (Westport, CT: Greenwood, 1985), p. 32.
[15] *Discourse on Method,* pt. VI.
[16] *Principia,* bk. II.
[17] C. Truesdell, *Essays in the History of Mechanics* (New York: Springer, 1968), p. 91.
[18] *Encyclopedia of Physics,* vol. III/3 (New York: Springer, 1965), p. 2.
[19] For a survey of problems and a sketch of possible solutions, cf. Hans Primas, *Chemistry, Quantum Mechanics and Reductionism* (New York: Springer, 1981). The book also contains a detailed discussion of the relation between chemistry and basic physics.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

biology, and social studies are even further removed from unification. Thus, instead of a multitude of particulars firmly tied to a set of time–invariant basic laws, we have a variety of approaches with unifying principles looming indistinctly in the background—a situation quite similar to what happened in Greece after Zeus's victory over the Titans.[20]

Thirdly, we are far from possessing a single consistent set of fundamental laws. Basic physics, the alleged root of all reductions, is still divided into at least two principal domains: the world of the very large, tamed by Einstein's general relativity, and the quantum world which itself is not yet completely united. "Nature likes to be compartmentalised," wrote Dyson describing this situation.[21] "Subjective" elements such as feelings and sensations which form a further "compartment" are excluded from the natural sciences, though they play a role in their acquisition and control. This means that the (unsolved) mind-body problem affects the very foundation of scientific research. Science has large lacunae; its alleged unity and comprehensiveness is not a fact but a (metaphysical) assumption, and those of its projections which work come from isolated areas and thus lack the destructive power ascribed to them. They show how certain sections of the world respond to crude approaches; they give us no clue about the structure of the world as a whole.

Finally, as the most fundamental and most highly confirmed theory of present day physics, the quantum theory rejects unconditional projections and makes existence depend on special historically determined circumstances. Molecules, for example, the basic entities of chemistry and molecular biology, do not simply *exist*—period—they *appear* only under well-defined and rather complex conditions.

If one still insists that the bits and pieces of science that are flying around today are superior by far to the analogous collections of a past age—a live nature, whimsical gods, etc.—then I must refer back to what I said earlier: the superiority is the result of having followed a path of least resistance. Gods cannot be captured by experiment, matter can. This point, incidentally, plays a role also

[20] Hesiod, *Theogony*, 820ff.

[21] *Disturbing the Universe* (New York: Harper & Row, 1979), p. 63. Superstring theoreticians try to overcome the remaining multiplicity, but all they have so far succeeded in doing is devising a language in which they can talk about everything without ever arriving at any concrete results. In the words of Richard Feynman: "I do feel strongly that this is nonsense" [P. C. W. Davies and J. Brown, eds., *Superstrings* (New York: Cambridge, 1988), p. 194]. Besides, even if superstring theoreticians succeeded in uniting basic physics, they would still have to deal with chemistry, biology, consciousness, etc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

within the sciences. "The great success of Cartesian method and the Cartesian view of nature," write R. Levins and R. C. Lewontin[22] commenting on the significance of the recent advances in molecular biology,

> is in part a result of a historical path of least resistance. Those problems that yield to the attack are pursued most vigorously, precisely because the method works there. Other problems and other phenomena are left behind, walled off from understanding by the commitment to Cartesianism. The harder problems are not tackled, if for no other reason than that brilliant scientific careers are not built on persistent failure. So the problems of understanding embryonic and psychic development and the structure and function of the central nervous system remain in much the same unsatisfactory state they were fifty years ago, while molecular biologists go from triumph to triumph in describing and manipulating genes (*ibid.*, pp. 2f.).

E. Chargaff[23] writes as follows:

> The insufficiency of all biological experimentation, when confronted with the vastness of life is often considered to be redeemed by recourse to a firm methodology. But definite procedures presuppose highly limited objects (*ibid.*, p. 170).

The insufficiency of science vis-à-vis the gods could not be expressed more clearly.

IV. DOGMATISM, INSTRUMENTALISM, RELATIVISM

There exist various ways of dealing with this situation. One is to disregard it and to continue describing the world in accordance with one's own pet metaphysics. This is the attitude of most scientists and scientific philosophers. It is a sensible attitude; it was the attitude of the educated Greeks and Romans who retained their gods amidst a flurry of philosophical objections. It does not solve our problem.

Instrumentalists react by dropping the second assumption. They do not drop it absolutely ("nothing exists"), however, but only with respect to certain entities.[24] A confrontation with alternative ontologies revives the problem.

Relativists accept the first assumption but relativize the second: atoms exist *given* the conceptual framework that projects them. The trouble here is that traditions not only have no well-defined bound-

[22] *The Dialectical Biologist* (Cambridge: MIT, 1985).
[23] *Heraclitean Fire* (New York: Rockefeller UP, 1978).
[24] This was realized by Duhem, who described a certain stage of the debate between realism and instrumentalism in astronomy as a battle "between two realist positions." *To Save the Phenomena* (Chicago: University Press, 1969), p. 106.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

aries, but contain ambiguities and methods of change which enable their members to think and act as if no boundaries existed: potentially every tradition is all traditions. Relativizing existence to a single "conceptual system" that is then closed off from the rest and presented in unambiguous detail mutilates real traditions and creates a chimera.[25] Paradoxically this is done by people who pride themselves of their tolerance toward all ways of life.

Relativists are right, however, when asserting that the temptation to project certain entities (gods, atoms) increases in some circumstances, and is diminished in others. Given the auspicious circumstances, the entities do indeed "appear" in a clear and decisive way. More recent developments in the interpretation of quantum mechanics suggest to regard such appearances as phenomena (Bohr's term) that transcend the dichotomy subjective/objective (which underlies the second assumption). They are "subjective," for they could not exist without the idiosyncratic conceptual and perceptual guidance of some point of view (which need not be available in explicit form); but they are also "objective": not all ways of thinking have results and not all perceptions are trustworthy. New terminology is needed to adapt our problem to this situation.

V. HUMANS AS SCULPTORS OF REALITY

According to the first assumption, our ways of thinking and speaking are products of idiosyncratic historical developments. Commonsense and science both conceal this situation. For example, they say (second assumption) that atoms existed long before they were found. This explains why the projection received a response, but overlooks that vastly different projections did not remain unanswered.

A better way of telling the story is the following. Scientists, being equipped with a complex organism and embedded in constantly changing physical and social surroundings, used ideas and actions (and, much later, equipment up to and including industrial complexes such as CERN) to *manufacture*, first, metaphysical atoms, then, crude physical atoms, and, finally, complex systems of elementary particles out of a material that did not contain these elements but could be shaped into them. Scientists, according to this account, are sculptors of reality—but sculptors in a special sense. They not merely *act causally* upon the world (though they do that, too, and they have to if they want to "discover" new entities); they also *create semantic conditions* engendering strong inferences from known effects to novel projections and, conversely, from the projections to

[25] I have described this aspect of relativism in ch. 10 of my book *Farewell to Reason* (London: Verso, 1987).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

testable effects. We have here the same dichotomy of descriptions which Bohr introduced in his analysis of the case of Einstein, Podolsky, and Rosen.[26] Every individual, group, and culture tries to arrive at an equilibrium between the entities it posits and leading beliefs, needs, expectations, and ways of arguing. The separability assumption arises in special cases (traditions, cultures); it is not a condition (to be) satisfied by all and it certainly is not a sound basis for epistemology. Altogether, the dichotomy subjective/objective and the corresponding dichotomy between descriptions and constructions is much too naive to guide our ideas about the nature and the implications of knowledge claims.

I do not assert that any combined causal-semantic action will lead to a well-articulated and livable world. The material humans (and, for that matter, also dogs and monkeys) face must be approached in the right way. It *offers resistance;* some constructions (some incipient cultures—cargo cults, for example) find no point of attack in it and simply collapse. On the other hand, *this material is more pliable than is commonly assumed.* Molding it in one way (history of technology leading up to a technologically streamlined environment and large research cities such as CERN), we get elementary particles; proceeding in another, we get a nature that is alive and full of gods. Even the "discovery" of America which I used to support the separability assumption allowed some leeway, as is shown by Edmondo O'Gorman's fascinating study, *The Invention of America.*[27] Science certainly is not the only source of reliable ontological information.

It is important to read these statements in the right way. They are not the sketch of a new theory of knowledge which explains the relation between humans and the world and provides a philosophical grounding for whatever discoveries are being made. Taking the historical character of knowledge seriously means rejecting any such attempt. We can describe the results we have obtained (though the description will always be fatally incomplete), we can comment on some similarities and differences that have come to our attention, and we can even try to explain what we found in the course of a

[26] Cf. the reprint in J. A. Wheeler and W. H. Zurek, eds., *Quantum Theory and Measurement* (Princeton: University Press, 1983), p. 42. The present essay is firmly based on Bohr's ideas. Reading the epilogue of Paul Hoyningen's *Die Wissenschaftsphilosophie Thomas Kuhns* (Braunschweig: Vieweg, 1989), I also realize that its ideas are very similar to, and almost identical with, Kuhn's as yet unpublished, later philosophy. I asked Hoyningen how he would explain this pre-established harmony (when writing my paper I was not familiar with Kuhn's later philosophy). His answer—"Reasonable people think along the same lines"—seems entirely acceptable.

[27] Bloomington: Indiana UP, 1961.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

particular approach "from the inside," i.e., using the practical and conceptual means provided by the approach (the theory of evolution, evolutionary epistemology, and modern cosmology belong in this category). We can tell many interesting *stories*. We cannot explain, however, how the chosen approach is related to the world and why it is successful, in terms of the world. This would mean knowing the results of all possible approaches or, what amounts to the same, we would know the history of the world before the world has come to an end.

And yet we cannot do without scientific know-how. Our world has been transformed by the material, spiritual, and intellectual impact of science and science-based technologies. Its reaction to the transformation (and a strange reaction it is!) is that we are stuck in a scientific environment. We need scientists, engineers, scientifically inclined philosophers, sociologists, etc., to deal with the consequences. My point is that these consequences are not grounded in an "objective" nature, but come from a complicated interplay between an unknown and relatively pliable material and researchers who affect and are affected and changed by the material which, after all, is the material from which they have been shaped. It is not therefore easier to remove the results. The "subjective" side of knowledge, being inextricably intertwined with its material manifestations, cannot be just blown away. Far from merely stating what is already there, it created conditions of existence, a world corresponding to these conditions and a life that is adapted to this world; all three now support or "establish" the conjectures that led to them. Still, a look at history shows that this world is not a static world populated by thinking (and publishing) ants who, crawling all over its crevices, gradually discover its features without affecting them in any way. It is a dynamical and multifaceted being which influences and reflects the activity of its explorers. It was once full of gods; it then became a drab material world; and it can be changed again, if its inhabitants have the determination, the intelligence, and the heart to take the necessary steps.

PAUL FEYERABEND

University of California/Berkeley

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

COMPUTATION AND CONSCIOUSNESS*

> Their favorite target was the rigid and silent Olympia, who, her beautiful appearance notwithstanding, was assumed to be hopelessly stupid, which was thought to be the reason Spalanzani had kept her so long concealed. Nathanael heard all this, not without inner fury, but he said nothing. "What would be the use," he thought, "of proving to these fellows that it was their own stupidity which precluded them from appreciating Olympia's profound and beautiful mind."
>
> from E. T. A. Hoffmann's *The Sandman*

THERE is no more amazing and puzzling fact than that of consciousness. Little wonder, then, that Philosophia, daughter of Thaumas, should wile away so many hours speculating on the nature of the mind. For, as Thomas Nagel has noted, it is consciousness that makes the mind-body problem intractable, it is the unfathomable gap between physical process and subjective awareness which mocks our search for the filaments that bind the corporeal and the mental together. Any program that holds promise for solving the mind-body problem deserves our closest attention. I wish to focus on one such program in this essay.

There is little doubt that humans have a mental life, because we have brains. Yet granted that brain activity somehow supports mental activity, the question still arises: In virtue of what do brains support minds? Which properties of our gray matter are essential, and which accidental, for our mentality? What level of description or abstraction distills the essence of mental life? The answer to this question is the beginning of wisdom in the philosophy of mind.[1]

* An embryonic version of this paper was presented at the University of Pittsburgh in January 1988. I am particularly indebted to Kevin Kelly, Clark Glymour, and Peter Spirtes for forcing me to clarify the main points at issue. I do not know that any of them would endorse any part of the argument. I am sure that they would all still find Olympia to be unpleasantly baroque, but I am afraid that, like poor Nathanael, I have fallen in love with her.

[1] One might wonder why the answer is not also the end of wisdom in the philosophy of mind. In this paper, I mean only to consider that part of a solution of the mind-body problem which purports to detail necessary and sufficient conditions for physical systems to have minds. It might be possible to discover such conditions and still be puzzled about how subjective states arise from certain physical conditions. Making the connection between mind and body intelligible, seeing how subjective states can be brought about by objective conditions, is a further problem, and of another order of magnitude.

0022-362X/89/8608/407–432
© 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

One clear answer to this question is now being widely championed: the *computational* structure of the brain is what bestows mental properties. We must abstract away the particular physical, biochemical, and neural features of brains to see what really makes them tick. The computational or information-theoretic description captures what is of importance.

The computational approach has been most intensively applied to problems of intentionality, language use, and representation. But, if it is to untie the Gordian knot uniting *res cogitans* with *res extensa,* the computational paradigm must eventually directly face the problem of consciousness. So we must ask: Is a computational theory of consciousness possible? Or, given the centrality of consciousness to the mind-body problem, we may equally phrase the question: Is a computational theory of mind possible?

Before considering the implications of a computational theory of consciousness, let me pause to define our terms more clearly. By 'consciousness' I mean *subjective phenomenal states* or *modes of awareness.* The most obvious examples of conscious episodes are sensory: tickles, pains, visual experiences, and so on. But they are not confined to straightforwardly sensory events. There is a certain phenomenology associated, for example, with my pondering the fact that ice is made of water. In such cases, the phenomenal properties of the experience need not determine the content of the proposition entertained. If a causal theory of reference is correct, a molecule-for-molecule identical replica of my brain, if just brought into existence, may not be capable of entertaining the proposition that ice is made of water. Still, our best guess is that such a brain would support identical states of consciousness to mine, identical phenomenal states. In Nagel's terms, what it would be like to be the person with that brain would be just what it is like to be me. The content and structure of those subjective states of awareness are what I mean by 'consciousness'.

Of course, the thesis that physically identical brains would support phenomenally identical states of consciousness is not analytic. But some such physicalist assumption underlies all contemporary research into perception and neuro-physiology. Furthermore, it seems to be an essential thesis for the computationalist. For computational structure supervenes on physical structure, so physically identical brains are also computationally identical. Hence, any mental property that can be given a purely computational analysis ought to be shared by physically identical brains.

In the sequel, a somewhat stronger claim about supervenience shall be employed. States of awareness and sensory events take place

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

in time; they are fairly precisely datable. One can assert that Sam had a toothache at 12:05 or that Sheila spent five minutes wondering about Fermat's last theorem. A natural, indeed nearly inescapable, explanation for this is that conscious events and episodes supervene on concurrent physical events and processes. One's phenomenal state at a time is determined entirely by one's brain activity at that time. Hence, two physical systems engaged in precisely the same physical activity through a time will support the same modes of consciousness (if any) through that time. Let us call this the *supervenience thesis.*

Again, this is a substantive thesis, albeit one for which we have some fairly direct evidence. We know that we can induce pains and visual experiences at a time by stimulating particular areas of the brain at that time. We know that particular types of mental activity are directly correlated with concurrent types a brain activity as revealed, for example, by the electro-encephalograph. It is not a far leap to suppose that all modes of subjective awareness supervene on that brain activity.

It may be useful to intimate exactly how the supervenience thesis will be used in the sequel. Suppose that a system exists whose activity through a period of time supports a mode of consciousness, e.g., a tickle or a visual sensum. The supervenience thesis tells us that, if we introduce into the vicinity of the system an entirely inert object that has absolutely no causal or physical interaction with the system, then the same activity will still support the same mode of consciousness. Or again, if the activity of a system supports no consciousness, the introduction of such an inert and causally unconnected object will not bring any phenomenal state about.

The plausibility of this particular application of the supervenience thesis derives from the unity and completeness of occurent phenomenal states. A pain has its entire existence and being in the event of its being perceived, and its perception is a single, unified occurrence. This suggests that the supervenience space of the pain must also have a sort of unity or connectedness, presumably a causal connectedness. If an active physical system supports a phenomenal state, how could the presence or absence of a causally disconnected object effect that state? How could the object enhance or impede or alter or destroy the phenomenal state except via some causal interaction with the system? Since the phenomenal state is entirely realized at the time of the experience, only the activity of the system at that time should be relevant to its production. The presence or absence of causally isolated objects could not be relevant. This is all the supervenience thesis needs to say.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Having stated both what I mean by 'consciousness' and the supervenience relation that I take to hold for it, let us now turn to 'computational'. What is a computational theory of consciousness?

To avoid the quagmire of behaviorism, one must posit that the (actual and counterfactual) input-output relations of the brain do not alone determine its mental properties. The internal structure of the brain is essential: one must examine how it generates the output from the input. Broadly speaking, any program that is concerned with characterizing the internal structure of this brain activity may be denominated a form of *functionalism*. Functionalists examine how the brain is organized, how it does what it does.

If functionalism is so broadly defined, though, it becomes nearly vacuous. One of the "ways" that the brain does what it does is by being a neuro-physiological object, by transmitting nerve impulses, etc. One of the ways the brain is organized is its chemical organization, and its structure can be defined at the level of physical structure. If the level of organization or degree of abstractness of a functional theory of mind is not constrained, then it is bound to succeed, at least in giving sufficient conditions for consciousness. This will be guaranteed if physicalism is true. To give functionalism any real bite, then, some precise specification of the appropriate level of functional organization must be set. One exact solution to this problem is proposed by the computational or information-processing theories of mind. They posit that the important level of functional organization of the brain is its computational structure.

The fundamental notion of a computational description is that of a machine table. The table describes how the internal states of the machine are subjunctively connected to one another and to various input and output mechanisms. In the case of a Turing machine, the example we shall be concerned with, the machine table specifies how the machine in a given state would respond to an entry on the machine's tape by (a) changing the entry on the tape, (b) moving to another address on the tape, and (c) going into a new internal state. The machine table thus determines exactly how the state of the machine and tape will evolve given any input and initial state.

The foremost advantage of a computational description of a system is its precision. Once the internal states and proper operating parameters of a physical system have been specified, one can determine by straightforward physical analysis how the states would be subjunctively connected and hence what the appropriate machine table is. One can compare the machine tables of any two systems to see if they are the same, and thereby determine whether they have the same program or are, at a given time, performing the same

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

computation. The computational approach gives a determinate content to functionalism by precisely specifying the level of abstraction at which a physical system should be analyzed.

We should remark just how high a level of abstraction a computational description employs. The machine table does not directly specify what sort of physical or material properties a system running a program must have. Nor can much in the way of physical constraints be even indirectly derived from the machine table. We may contrast, by way of example, the functional description of a valve lifter. Jerry Fodor[2] has used the concept of a valve lifter as an archetypical functional concept. From the fact the something is a valve lifter (as opposed to, say, a camshaft), one cannot infer its precise material structure or composition. To be a valve lifter is to play a certain functional role in the internal economy of an engine. Still, there seem to be some physical constraints on valve lifters: they must, for example, transmit enough energy to lift a valve. A weak photon cannot be a valve lifter or (to use another functional type) a mousetrap, because it cannot provide the physical wherewithal to lift a valve or catch a mouse. These physical constraints derive from the fact that the functions involved are ultimately specified by reference to physical objects and properties. Engines must produce mechanical work, mousetraps must immobilize a given type of animal. The items playing certain functional roles in service of these ends may have to meet some physical standards in order to achieve them. So, although these functional concepts can abstract from the exact specification of material structures, still some fairly strong physical capacities may be implied by them.

In contrast, the terms of a computational description involve no ultimate physical or mechanical goal to be achieved. The only physical requirements that a system must meet in order to instantiate a certain machine table are that (1) there must be at least as many physically distinguishable states of the system as there are machine states in the table, (2) the system must be capable of reacting to and changing the state of the tape, and (3) there must be enough physical structure to support the subjunctive connections specified in the table. Similar remarks apply to the tape itself.[3]

[2] *Psychological Explanation* (New York: Random House, 1968), pp. 113 ff.

[3] If brains support mental states in virtue of their computational structure, then it is clear that minds could be realized in almost any stable material substrate. The intuition that persons may only accidentally be made of flesh and blood dates back at least to Socrates the Younger (cf. Aristotle, *Metaphysics* 1036b20 ff.), and is part of the credo of the AI community. Witness the strong reaction to John Searle's suggestion that "whatever else intentionality is, it is a biological phenomenon, and it

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

A computational theory of consciousness must therefore hold that particular physical systems are capable of consciousness because they can be correctly described as machines that have been appropriately programmed. That is, there must be some identification of internal states and parameters of normal operation such that, under that interpretation, the machine instantiates an appropriate machine table, and any system that satisfies that table will be capable of consciousness. Of course, there will presumably be an infinite number of such programs that would bestow the capacity for consciousness.

Furthermore, if we move from the capacity for consciousness to its exercise, a computational theory of mind must hold that a system actually is conscious, is supporting an occurrent mode of awareness, when it is actually running the appropriate program on the appropriate tape. So, if for the last five minutes I have had a toothache, or the subjective experience of pondering Fermat's last theorem, then there must be some program and some state of the machine tape such that anything that runs that program on that tape (starting in a particular machine state and progressing through a finite number of cycles) will create or support or underlie or be associated with an occurrence of that mode of consciousness. Finally, a computational theory of consciousness must hold that a necessary condition for supporting consciousness be that a system be describable as a nontrivial computational system performing a nontrivial computation. Otherwise, a system with no interesting computational structure could be conscious, so computation could not be essential to consciousness.

If we cast these last two requirements in terms of Turing machines,[4] we have the following two principles. Take any type of phenomenal conscious state ϕ:

Sufficiency Condition: There must be some program π and some state of the machine tape τ and some sequence of machine states $S_{[0]}$, $S_{[1]}$,

---

is as likely to be as causally dependent on the specific biochemistry of its origins as lactation, photosynthesis, or any other biological phenomenon" ("Minds, Brains and Programs," *The Behavioral and Brain Sciences III*, 3 (1980): 424). Zenon Pylyshyn, for example, asserts that "we cannot take as sacred anyone's intuitions about such things as whether another creature has intentionality—especially when such intuitions rest (as Searle's do, by his own admission) on knowing what the creature (or machine) is *made of*. . . . Clearly, intuitions based on nothing but such anthropocentric chauvinism cannot form the foundation of a science of cognition" (*ibid.*, p. 443). As we will see, there are other levels of description less abstract than the computational level which also sustain the possibility that the functional organization can be realized in almost any material.

[4] Similar conditions can easily be stated for finite-state machines. I take the Turing machine as the most general case.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

$\cdots S_{[N]}$ such that any machine programmed with π, operating on τ, and performing the computation $S_{[0]}$, $S_{[1]}$, $\cdots S_{[N]}$ (in accordance with π) will, during the time that the computation is taking place, support the mode of consciousness ϕ.

Necessity Condition: A necessary condition for a physical system to support ϕ is that it be describable as a Turing machine performing some nontrivial computation.

(The subscripted number in square brackets denotes the ith step in this particular computation. So this computation starts in state $S_{[0]}$ and with the read/write head at location $T_{[0]}$, next moves to $S_{[1]}$ and $T_{[1]}$, and so on. Subscripts without brackets indicate a general numbering of the different machine states and tape addresses. So, if in the course of the computation, the second tape address is visited at the 5th and 18th steps, $T_{[5]} = T_{[18]} = T_{[2]}$.)

The burden of the remainder of this essay will be to demonstrate that the sufficiency condition, the necessity condition, and the supervenience thesis form an inconsistent triad, and hence that an acceptable computational theory of consciousness is not possible. Let me first, however, try to forestall some possible misunderstandings.

I. WHAT THIS ARGUMENT IS NOT

In developing the argument that follows, I shall construct a machine that will run the program π. The machine will be somewhat peculiar and complicated, but its purpose should not be misconstrued. The point of the machine is to demonstrate the inconsistency of the three conditions listed above. This line of argument must be distinguished from another common dialectical gambit which may be called the *ploy of funny instantiation.*

The ploy of funny instantiation, in its crudest form, operates by first pointing out that any Turing machine can be realized in some peculiar materials (e.g., water pipes or toilet paper or windmills and old beer cans), and then appealing to intuitions that that sort of stuff just cannot be conscious or have mental states or intensionality. I fully reject this ploy, because I approach the mind-body problem as one thoroughly mystified by the relationship between physical activity and consciousness. Although I may have intuitions that windmills and old beer cans, no matter how cleverly connected, cannot be conscious, still I cannot think of one reason to accord those intuitions any weight. How pulses of water in pipes might give rise to toothaches is indeed entirely incomprehensible, but no less so than how electro-chemical impulses along neurons can. For all I know, a conscious state may equally well be created by either of these sorts of process.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

John Searle's Chinese room argument is a sort of funny instantiation ploy, but with a twist.[5] Searle does not directly argue that this particular system, a human shuffling some cards about, is not the kind of thing which can be conscious or have intensionality or understanding. Rather, since the man is just the kind of thing which is conscious, Searle simply presupposes that whatever intensionality or understanding arises by shuffling the cards would have to belong to the same mental entity, the same self, that is associated with the brain activity of the man in the room. Since *that* personality does not come to understand Chinese, Searle assumes that nobody or nothing understands Chinese. But the inference does not go through.

The problem is perhaps clearer in the case of sensations. Suppose that someone claims that the operation of a particular program will always create or support the experience of a toothache. If we allow the man in the toothache room to be our Turing machine and perform the appropriate computational activity, we do not suppose that thereby his tooth would start to hurt. Rather, a phenomenal state would supervene on his activities, a phenomenal state disjoint from his consciousness. For, on this theory, his consciousness is created by the pattern of activity of his neurons, not by the computational structure of the manipulation of the cards. There is no reason to insist that the phenomenal states associated with the former must combine with the phenomenal states associated with the latter into one self.

This is exactly the solution to Searle's problem offered by several of the commentators on his paper. Robert Wilensky writes: "Searle's mistake of identifying the experiences of one system with those of its implementing system is one philosophers often make when referring to AI systems" (*ibid.*, p. 450). Aaron Sloman and Monica Croucher suggest that an appropriate analogy to the relationship between the consciousness of the man in the room and the consciousness created by the man in the room may be found in cases of multiple personalities (*ibid.*, p. 448). And Marvin Minsky directly asserts that "in the case of a mind so split into two parts that one merely executes some causal housekeeping for the other, I should suppose that each part —the Chinese rule computer and its host—would then have its own separate phenomenologies—perhaps along different time scales" (*ibid.*, p. 440). (Note that, although Searle's argument is concerned with intensionality and understanding, these writers also speak of consciousness and phenomenologies, implying that they accept a computational theory of consciousness.) Searle has done nothing

[5] The Chinese room was constructed in Searle, *op. cit.*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

to discount the possibility of simultaneously existing disjoint mentalities.

The time-scale problem underlies another sort of funny instantiation ploy. The clock of a computer can be slowed down indefinitely without affecting its computational structure. A machine might, for example, go through only one cycle every 1,000 years. It might take billions of years to complete the computations from $S_{[0]}$ to $S_{[N]}$. It might spend 999 years out of each 1,000 disassembled, parts widely scattered, only being put together for the single millennial cycle. None of these facts would affect its computational description, so, if we accept the sufficiency condition, there would still be associated with the billion year process a conscious episode of type ϕ. This does seem exceedingly odd, but I do not see any way of promoting this oddness into a rational objection to the sufficiency condition.

I am not, then, constructing this machine in order to appeal to primitive intuitions about what sort of materials can or cannot support consciousness. The question I want to ask is rather the following: Given that consciousness supervenes on physical processes and activity, what is the minimum level of physical activity that a machine must perform in order to run program π from $S_{[0]}$ to $S_{[N]}$, and hence (according to the computational theory) to support phenomenal state ϕ? What is the laziest machine that can be conscious? This problem shall guide our construction.

II. THE RULES OF THE GAME

In order to run π from $S_{[0]}$ to $S_{[N]}$, the following three conditions must be fulfilled. First, the machine itself must pass sequentially through N machine states. Second, the read/write head of the machine must sequentially visit tape locations $T_{[0]}$ to $T_{[N]}$. In doing so, it may have to change the data entries of any or all locations. These first two conditions contain all of the activity that can be explicitly derived from the definitional conditions of what it is to run π on τ. But there are other conditions that may somehow implicitly demand that much more complicated activity is needed. For, thirdly, the physical state of the machine must also support all of the counterfactuals inherent in the machine table. Not only must the machine actually go from $S_{[0]}$ to $S_{[1]}$ given that the data location $T_{[0]}$ actually contains, say, a 0, but it must also be constructed so that it would have gone into $S'_{[2]}$ (in accordance with π) had $T_{[0]}$ contained a 1. And the counterfactuals ramify geometrically: for every state it might have gone into, the program will specify where it would have gone after that (for each possible input), and so on. Thus, even though the machine may go into only a small number of states in running π from $S_{[0]}$ to $S_{[N]}$, still in order to be running π it may have to have a tremendous number of

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

states accessible to it, all interconnected in exactly the right way. This at least prima facie suggests that the physical activity needed to run π must be immensely more complicated, and the transitions from one state to another immensely more complex, than the two explicit conditions on physical activity suggest. Our machine will show that this plausible argument is not correct: the physical activity involved in running a program may be limited to the most trivial means of fulfilling the explicit conditions, with the satisfaction of the third condition on counterfactual structure requiring almost no additional activity at all.

In constructing our lazy machine, certain tricks shall be ruled out of order. For example, in imagining how little physical activity is needed to pass from some state S_i to another one S_j, someone might suggest that no activity is needed. Let a rock sitting on a table be the machine. Now let S_i be: sitting on the table from 12:00 to 12:01. Let S_j be: sitting on the table from 12:01 to 12:02. The machine will effect a transition between the two states without undergoing any physical change at all. I shall take such tricks to be inadmissible; it is in the spirit of computationalism that the machine states be physically meaningful states and that the state transitions be physical activities. I shall also require that the tape addresses be physically distinct entities and that the informational state of an address (0 or 1) be specifiable in terms of its intrinsic physical state. Lastly, the state of the machine must be defined without reference to the state of the tape, and vice versa. Even with these restrictions, we shall see that any program can be run with an uncomfortably small amount of physical activity. Without them, the activity could be reduced even further.

III. THE CONSTRUCTION

In constructing our machine, I first posit that we already have a machine programmed to run π. Let us call this machine *Klara*. Klara can in principle be constructed in whatever manner and out of whatever materials one prefers. She may be silicon or protein, electrochemical or hydraulic. I shall imagine her as a clockwork mechanism, but nothing turns on this choice. Klara may well be unimaginably complex; as a clockwork she may have to be as large as the solar system. These details will ultimately be irrelevant to the construction.

Although Klara's own composition will be subject to no constraints, Klara's tape is to have a very particular form. She will store and retrieve information in a (potentially) infinite series of water troughs. Each trough can store one binary bit of information by being either full of water or empty. We can imagine Klara's read/write head getting information from an address by dropping a float

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

down into the trough to determine whether it is full or not. Entries can be changed by either filling an empty trough or opening a drain on a full one. To begin with, then, a section of Klara's tape would be as depicted in figure 1.

Now, certain sorts of manipulation can be performed on Klara's tape which (with compensating adjustments) would in no way affect her computational structure. We may, firstly, arbitrarily alter the order of any finite number of the troughs, so long as we provide Klara with a map of the alterations. Before this change, if Klara is at location 18 and has been instructed to go to the location +2, she will just move two troughs to the right. After the rearrangement, she would first determine that she is at address 18, calculate that she must go to address 20, locate that trough on the map, and move her read/write head there. So the spatial sequence of the troughs need not reflect their "computational sequence." We may so contrive that any sequence of addresses lie next to each other spatially.

Secondly, any tape address can be *bi-located.* If it is convenient for some reason to have a single address be accessible from two different spatial locations, we need only set up a trough at each location and connect them with a pipe. The two now jointly form one address, with its informational content available from each location. The address can also have its contents changed, i.e., be filled or emptied, by means of an operation performed at either trough. Again, information about such bi-location (or more highly multiple location) can be encoded in Klara's map.

Aside from these possible manipulations of the tape, we need introduce only one more item. A *block* is a device that can halt the operation of Klara and freeze her in a particular machine state. If she is clockwork, we can imagine placing a simple block of wood between

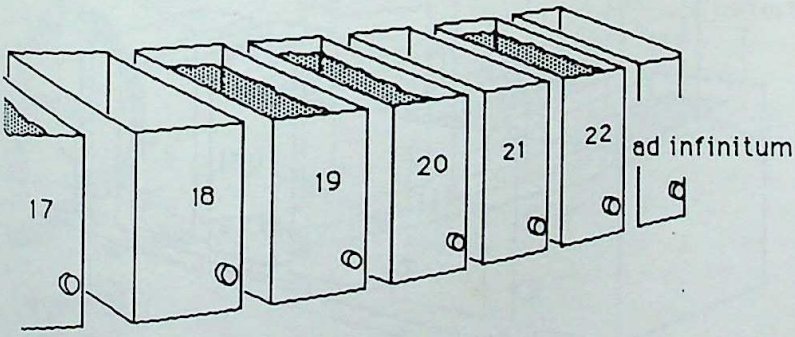


Figure 1: Klara's Original Tape

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

her main operating gears, immobilizing them. The gear teeth of a blocked machine may well not even physically contact each other, remaining suspended motionless in space. Unblocking a machine allows it to continue running as normal.

The first step in our construction is to rearrange Klara's tape so that addresses $T_{[0]}$ to $T_{[N]}$ lie spatially in sequence, $T_{[0]}$ next to $T_{[1]}$ next to $T_{[2]}$, etc. (Recall that the numbering in square brackets, $T_{[n]}$, indicates the trough visited at step n of the computation.) Given the possibility of arbitrarily reordering and bi-locating troughs, this must be possible. We start with $T_{[0]}$, move $T_{[1]}$ next to it, and so on. If $T_{[i]} = T_{[j]}$ for some $i < j$, so the program requires the machine to "go back" to an address already visited, we simply bi-locate the address by connecting two troughs with a pipe. Any address can be multiply located to any finite degree. After the rearrangement, the relevant section of tape would look as depicted in figure 2.

Now, when Klara operates π on τ from $S_{[0]}$ to $S_{[N]}$ her read/write head will just sequentially move to the right. This may be accomplished by a very complex operation, by scanning address numbers and performing calculations and referring to maps, but the net result will be that Klara will always move to the next trough along, read it, perhaps fill or empty it, move to the next, etc., until reaching $T_{[N]}$.

We will begin constructing our new machine, Olympia, by contriving the simplest means of performing this sequence of operations on the tape. The task is nearly trivial: we need only an armature designed to travel left to right which will fill or empty the appropriate troughs as it passes by. Emptying might be accomplished by hitting a rod attached to the drain plug of the trough. Filling can be accomplished by a hose attached to the armature. If the trough being passed is to be filled, the water is allowed to flow in. If not, a barrier

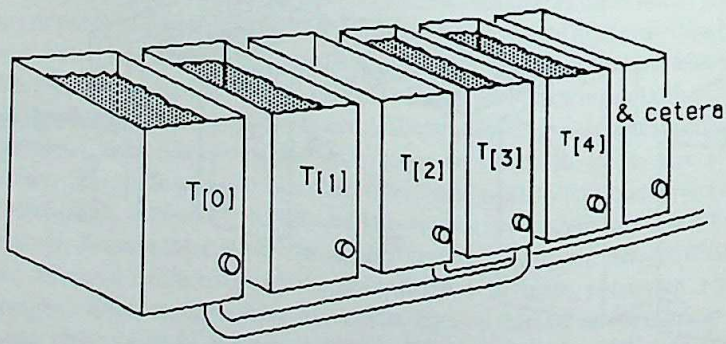


Figure 2: The Reordered Tape

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

placed at the top of the trough might divert the water away. So far, then, Olympia is just the device of figure 3.

As depicted, Olympia's armature would strike the first rod and empty $T_{[0]}$, leave $T_{[1]}$ unchanged, fill $T_{[2]}$ and $T_{[3]}$ ($=T_{[0]}$, and so at that point empty), etc. So she can already fulfill one of the requirements of a machine operating π on τ: the manipulations of the tape will be done correctly and in the appropriate order. The armature also is doing enough to fulfill a second requirement: the armature passes sequentially through N distinct physical states. One need only define as a distinct state the state of the armature being over a particular trough. If we could identify the state of being over the trough marked '$T_{[i]}$' with the machine state $S_{[i]}$ (for $i = 0$ to N), then Olympia would already be going through the appropriate sequence of machine states as the armature moves, and would already be running π. Of course, one cannot simply by fiat command that the armature position constitutes the appropriate machine state. A particular physical state only becomes interpretable as a machine state of a system programmed with π in virtue of standing in the right counterfactual or subjunctive relations to the tape and to the whole constellation of other states in the machine table for π.

Indeed, as matters stand, it seems impossible to maintain that the armature alone is instantiating any program or is doing any computing at all. Its operations are entirely oblivious to the state of the tape. The sequence of states it passes through and the manipulations it performs are entirely unresponsive to the data stored in the addresses. If one wants to ascribe a computation to it, the computation

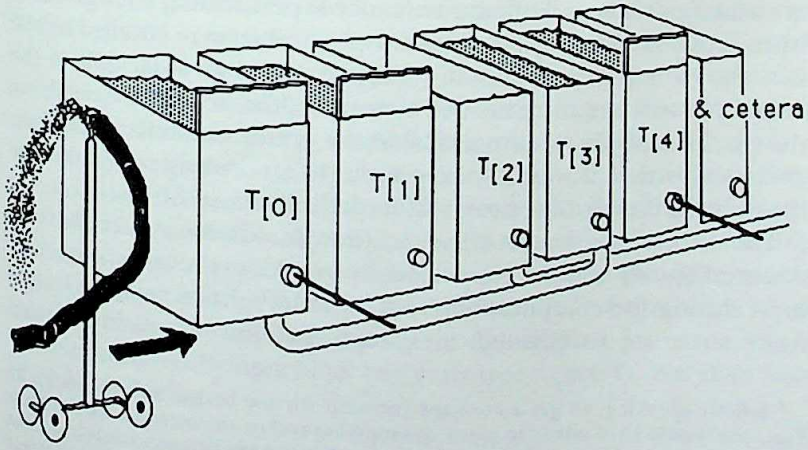


Figure 3: The Armature

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

would be just a constant function that gives the same output for any input.[6] Such a program would be, in as strong a sense as possible, trivial. It would make a mockery of the computational theory of mind, if the "computation" of a constant function after the manner of the armature could be sufficient to support any mode of consciousness. Indeed, the necessity condition asserts that any conscious entity must be describable as a nontrivial Turing machine running a nontrivial program. If one is to elucidate mental capacities as computational, the computations involved must not be just constant functions, especially constant functions "computed" in such a nonresponsive way. The computationalist must assert, then, that the armature alone, traveling from left to right, does not support a state of consciousness.

The point, again, is not that the armature and tape are the wrong kind of stuff to be conscious. For all I know, appropriately constructed troughs and water hoses could have toothaches, or think about Fermat's last theorem. The point is rather that the armature does not have the right kind of computational or information-processing structure to be conscious according to the computationalist's own thesis. For the armature is not processing any information at all.

As yet, the computationalist has no grounds to be alarmed that the activity of the armature supports no conscious state. The computationalist thesis is committed to maintaining that any system that runs π on τ from $S_{[0]}$ to $S_{[N]}$ is conscious, but the armature clearly does not instantiate π. The vast majority of machine states needed to describe π have not been correlated with any possible physical state of the armature, nor are the counterfactuals implied by π supported by the armature's structure. Still, the armature is performing enough activity to satisfy the explicit demands for physical change entailed by the fact that a machine is running π on τ. All that is lacking is the counterfactual structure needed to support the program. Somehow, the machine needs to have available the entire constellation of machine states that it would have gone into (according to π) had the data entries in the tape been different from those of τ.

The crux of the matter appears, though, when we note that the counterfactuals can be supported by machinery that is physically inert during the computation. That is, although we must add some more structure to Olympia to ensure that she is running π, that

[6] Strictly speaking, to get a constant function for the output in troughs $T_{[0]}$ to $T_{[N]}$, one would have either to place an emptying rod or to remove a barrier from every trough location, depending on whether the trough is respectively empty or full after the associated machine state $S_{[i]}$ has been completed. This addition would have no effect on the construction.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

structure need not do anything, need not be physically active, while this bit of π is running. We already have all the physical activity needed to run π.

So long as the troughs are filled according to the specification τ, the armature will perform all of the correct manipulations on it. What we need is some extra structure that would be activated were any of the tape addresses to hold different data. We can achieve this as follows.

At each trough location $T_{[i]}$ we attach a float. The float can be either up or down. Each float is held fixed in place by a latch, and the latch is released only when the armature passes by. Initially, each float is fixed in the position corresponding to the water level in $T_{[i]}$ when the machine goes into state $S_{[i]}$ during a normal run of π on τ. If during such a run address $T_{[i]}$ is full when the ith step is reached, we will fix the float up; if empty, we fix it down. So arranged, during a normal run of π on τ, none of the floats will in fact change position. When the latch is released the float is already set at the appropriate water level and so stays still.[7] Were the program to be run with the tape in a different state τ', in which the content of a trough $T_{[i]}$ differs from that in τ, then the float will either rise or fall when the armature reaches $T_{[i]}$ and releases the latch. So for any tape state that would cause a machine running π to evolve differently than it will running on τ, there will be at least one float that will move when the armature goes by.

The floats have provided Olympia with some subjunctive sensitivity to variations from τ. Now it is just a matter of hooking that sensitivity into the right machinery. To do this, we make $N+2$ copies of Klara. The first copy is blocked in the state $S_{[0]}$, and the read/write head is placed in trough $T_{[0]}$. The second copy is blocked in $S_{[1]}$ and set in $T_{[1]}$, and so on up to $S_{[N+1]}$. The block in the first machine (which is, recall, a small bit of wood jamming the gears) is attached by a chain to the float at $T_{[0]}$, and so on until the machine blocked in state $S_{[N]}$ is attached to the float a trough $T_{[N]}$. Now, if any float should rise or fall, it will unblock the corresponding copy of Klara. The machine blocked in state $S_{[N+1]}$ is set so that it will be set running, if the armature completes its trip from $T_{[0]}$ to $T_{[N]}$. Finally, a return chain ensures that as soon as any of the blocked machines is set running, the armature is immediately shut down.

We now have Olympia in all her glory (see figure 4). We need only give the definition:

[7] The latch is reengaged before the armature itself does anything to affect the level. Only the water level at the moment that the armature arrives at the trough determines the float's behavior.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

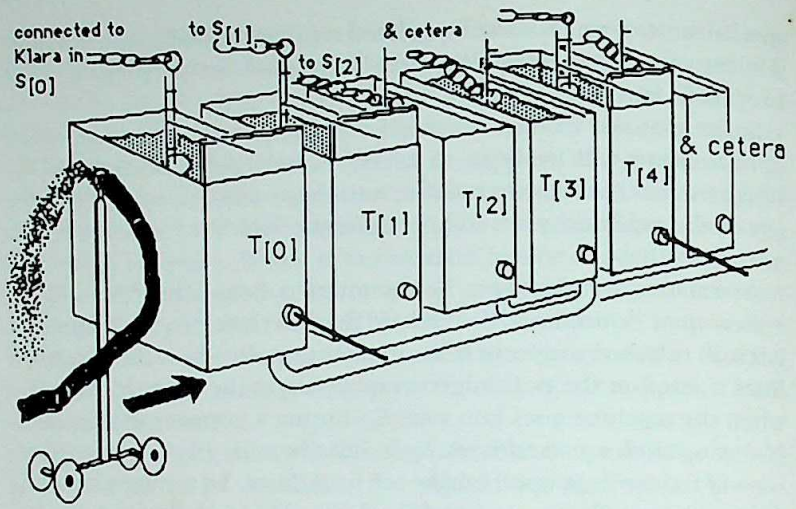


Figure 4: Olympia Unveiled

Olympia is in state S_i iff either one of the copies of Klara is unblocked and is in state S_i or $S_i = S_{[j]}$ for some j and the armature is running and passing over trough site $T_{[j]}$.

A little analysis confirms that Olympia is now programmed to run π. Once one of the copies of Klara gets unblocked, Olympia will run π because Klara is programmed to run π. But even while no copy of Klara is unblocked, while only the armature is moving, Olympia still instantiates π. For as they are now defined, the machine states are tied together by the right subjunctive relations. If we start Olympia running on τ she will, by construction and according to the definitions, pass through exactly the sequence of machine states and perform exactly the tape manipulations demanded by π. And, if we start Olympia running with the tape in a state other than τ, then at exactly the point that the armature alone would go wrong (where the data entry differs from τ), one of the copies of Klara wil be unblocked to take over the task. Thus, no matter what the state of the tape, the sequence of machine states as defined and the sequence of manipulations of the tape will be in exact accordance with π. But being programmed with π is nothing over and above having just these dispositions.

IV. THE BIND

We are now in a position to demonstrate the conflict that arises among our three theses. According to the sufficiency condition, when Olympia runs on τ and operates π from $S_{[0]}$ to $S_{[N]}$, she is

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

conscious and experiencing phenomenal state ϕ. But, when Olympia runs on τ, all that happens physically is that the armature tracks from left to right, emptying and filling troughs and momentarily unlatching floats. In particular, absolutely no physical activity occurs in any of the blocked copies of Klara. The floats, when released, all remain still. No physical or causal process need flow along the chains that rest, we may imagine, inert and with links not even in physical contact. Throughout this vast clockwork universe that we have appended to the armature an absolute stillness prevails; gears hang motionless in the air, not touching. But the supervenience thesis tells us that any mode of consciousness which occurs during the time of the computation must supervene on the physical activity and processes that occur through that period. Hence, Olympia's phenomenal state ϕ must supervene solely on the activity of the armature (and perhaps the floats), since that is all the physical activity there is. The masses of idle machinery form no part of the supervenience space, since nothing happens there. We have already seen, however, that, without the machinery, the armature alone cannot be interpreted as performing any nontrivial computation, and so, by the necessity condition, the conscious state ϕ cannot supervene on its activity alone.

In short, the computationalist is committed to the claims that the armature moving without the extra machinery hooked up cannot be conscious and that the system composed of the armature moving with the machinery hooked up must be conscious. But the physical activities that occur with and without the idle machinery connected are exactly identical, so these two claims contradict the supervenience thesis.

As has been already noted, the supervenience thesis implies that the presence or absence of inert, causally isolated objects cannot effect the presence or absence of phenomenal states associated with a system. In Olympia's case, the objects are huge, unwieldy, and unimaginably complex, but still inert, neither active in themselves not exerting any influence over the armature. So any existing phenomenal state does not supervene on them. But the supervenience space of Olympia's computational description, indeed whether she is computing at all, depends vitally on the counterfactuals that the idle machinery supports. Hence, her conscious phenomenal states cannot derive from her computational structure. A computational theory of consciousness is not possible.

Since a computationalist cannot give up the sufficiency condition, and since it would tear the guts out of the notion that mental states arise from a complex computational structure to give up the necessity condition, our central focus must be the plausibility and force of

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the supervenience thesis. Here two avenues of approach are possible. Either the computationalist can argue that the thesis is fine but that, contrary to appearances, the physical activities with and without the machinery hooked up are not the same, or else he can try to renounce the thesis altogether. Let us consider these possibilities in turn.

The supervenience thesis depends upon some independent notion of physical activity, and, in our application of it, upon the notion of sameness of physical activity. It would be beyond our needs to try to formulate an exact definition of physical activity here, but one line of argument deserves our attention. Perhaps one might urge that the counterfactuals supported in the first case differ from those supported in the second case, and *ipso facto* the physical processes occurring in the two instances must differ. Now, the argument by *ipso facto* is pleasingly short and crisp, but carries little conviction. We do have intuitions about sameness and difference of physical activity, and we cannot allow the computationalist just to build whatever is handy into the individuating conditions for types of physical process. Otherwise, even the Cartesian dualist can embrace the supervenience thesis: wherever there is a difference in the state of a thinking substance, *ipso facto* there is a difference in physical activity. But the notion that difference of counterfactual structure implies difference in physical process does not fare well when confronted with cases.

Consider, for example, a deterministic pin-ball game.[8] If we set the ball off at a specified place and with a specified velocity, it will always trace exactly the same path through the board. The only physical processes involved are those associated with the motion of the ball down the board and the interactions of it with the pins that it encounters. If we now remove a few pins, pins which the ball never touches on this path, pins which perhaps it never even comes near, the ball will continue to retrace exactly the same path in exactly the same fashion. The physicist's explanation of why it traces just that path will remain exactly the same. No motion or energy transfer or change of state occurs anywhere outside the path of the ball.[9] From a physical point of view, the processes and actions that occur with or without the peripheral pins in place are identical, for there are no physical processes occurring outside the path. But the counterfac-

[8] I owe this illustration to Kevin Kelly.

[9] Of course, at a subatomic level, there are all sorts of motion, but that is not essential to the case. The facts would remain the same in a Democritean world where the matter is just solid and totally inert stuff.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tuals in the two cases are different. For had the ball been given a different initial push, a push that would have carried it into the region where the pins are removed, then its path would be different in the two cases. So, although the counterfactuals supervene on the entire physical state of the system, differences in counterfactuals about the evolution of the system need not imply differences in physical processes that are evolving at a time.

To avoid getting bogged down in general claims, let us look more closely at the details of Olympia, and at the sorts of changes which are sufficient to defeat the support of counterfactuals provided by the blocked machinery. Above we considered the cases of Olympia running with and without the idle machinery present. Because of the immense quantity of machinery involved, one might misgive that its removal would necessitate some considerable change in the physical happenings associated with the machine. To alleviate such doubts, here are two cases in which the support can be neutralized by changes that can hardly be construed as altering the physical activities present.

An Argument by Addition: Suppose we run Olympia, fully connected, on τ so that (according to the supervenience thesis) the conscious state ϕ occurs. Now we reset her (and the tape) to run again, but we add a secondary block to each of the copies of Klara. The second block might be a thin piece of metal suspended between the frozen gear teeth. It need not be in physical contact with any part of the machinery. Now, however, were the first block to be removed (which will not, of course, happen when we run Olympia on τ from $S_{[0]}$ to $S_{[N]}$), the gears would contact the second block and jam. The copies of Klara no longer support the right counterfactuals, so on the second run Olympia is not conscious. But, given that the second blocks in fact never even touch any part of the machinery, exerting no physical influence or force at all, how could the physical activity taking place in Olympia during the first run be said to differ from that in the second? Speaking loosely, how could the rest of the system know that the blocks are even there?

An Argument by Subtraction: Suppose that, as we repeatedly run Olympia on τ, the chains connecting the floats to the blocks are slowly rusting. At first, associated with every run there exists a conscious state ϕ: Olympia feels, say, a toothache. Eventually, though, the chains so weaken that, were they to be pulled, they would break rather than unblock the machines. No unusual physical effect accompanies the passage of the point of critical weakness, nor does the progressive rusting cause any other alteration in the structure of the device. Still, according to the computationalist, once the critical point has been passed, suddenly no toothache accompanies the passage of the armature from left to right. Speaking loosely, how can the system know that the critical point has

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

been passed? How can the psycho-physical connection be broken by such a minor change in physical state which has no influence on the dynamical interactions occurring in the system?

The presence or absence of an inert second block that enters into absolutely no causal interaction with the system cannot plausibly be said to change the nature of the physical activities going on in the system. So the computationalist must renounce the supervenience thesis altogether. But even this does nothing to solve the puzzles presented above: even if mental states supervene on more than just the physical activity of a system, the crucial role of the entirely isolated block remains inexplicable. And, in countenancing the possibility of such effects, the computationalist would cut himself off from the research tradition from which the tradition grew. To see this, let us apply the point directly to brain activity.

The modern picture of brain function rests primarily on the notion of neural activity. The essential structure of mentation seems to be founded in patterns of neural firings. Because those firings can be analyzed as carrying information, the brain has come to be considered as an information processor. So let us suppose that some time in the future the electro-encephalograph is so perfected that it is capable of recording the firing of every single neuron in the brain. Suppose that researchers take two different surveys of a brain which match exactly: the very same neurons fire at exactly the same rate and in exactly the same pattern through a given period. They infer (as surely they should!) that the brain supported the same occurrent conscious state through the two periods. But the computationalist now must raise a doubt. Perhaps some synaptic connection has been severed in the interim. Not a synaptic connection of any of the neurons which actually fired during either period, or which was in any way involved in the activity recorded by the encephalograph. Still, such a change in connection will affect the counterfactuals true of the brain, and so can affect the subjective state of awareness. Indeed, the computationalist will have to maintain that perhaps the person in question was conscious through the first episode but not conscious at all through the second. I admit to a great degree of mystification about the connection between mind and body, but I see no reason to endorse such possibilities that directly contradict all that we do know about brain process and experience.

Whether the reason is enshrined in the supervenience thesis or not, our general picture of the relation between physical and mental reality firmly grounds the intuition that Olympia's experience cannot be changed by the presence or absence of the second set of blocks. These intuitions are not sacrosanct, but the computationalist especially abandons them at his own risk. For similar intuitions are often

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

appealed to in defending the appropriateness of computational analogies in the first place. One first step in arguments aimed at inducing assent to the possibility of computers that can think, or feel, or intend, is to imagine that some sort of prosthetie neuron made of silicon has been invented.[10] We are then to imagine slowly replacing some poor sap's brain bit by bit until at last we have a silicon brain that, our intuitions should inform us, can do all of the mental and intensional work of the original.

This is as yet a far cry from showing that anything has mental properties in virtue of its computational structure, but it is supposed to break down parochial species-chauvinistic views about there being any deep connection between mentality and organic chemistry. But the thought experiment rests on a tacit appeal to supervenience. How could it matter, one asks, whether the electrical impulses are carried by neurons or by doped silicon? The implication is that mentality supervenes only on the pattern of electrical or electrochemical activity. If the computationalist now must assert that the presence or absence of a piece of metal hanging untouched and inert in the midst of silent, frozen machinery can make the difference between being conscious and not, who knows what enormous changes in psychical state may result from replacing axons and dendrites with little copper wires? Should the computationalist reject the extremely general intuitions at play in assessing Olympia's case, no means of judging the plausibility or implausibility of any theory of mind seems to remain.

The silicon brain *Gedankenexperiment* is worthy of closer examination, as it turns on intuitions about supervenience similar to those we have invoked. But it requires a stronger supervenience thesis than ours, for the occurrent processes in the silicon brain are physically distinguishable from those in the organic brain in a way that the activities in Olympia with and without the extra blocks are not. Still, the suggestion that the pattern of electrical activity (however supported) determines the occurrent mental state has some plausibility, and allows us to abstract away from the organic features of the brain. Can we push the supervenience claim to an even higher level of abstraction?

The next obvious step is to suggest that the use of electricity (or electrochemical activity) is not essential to mentality. Let the ersatz neurons communicate by any causal process at all, and so long as the pattern of that activity is isomorphic to the brain, we still have consciousness and intensionality. At this stage of abstraction, we are

[10] Cf., for example, Pylyshyn's response to Searle, *op. cit.*, p. 442, or Clark Glymour's "Silicon Reflections" (typescript).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

allowed to replace axons and dendrites with, e.g., pipes, and to let pulses of water play the functional role of electrochemical discharges. This allows a greatly broadened class of materials out of which a mind can be built. The appropriate supervenience thesis at this level of abstraction would assert that mental events supervene on the pattern of causal interaction in the brain, where the exact physical or chemical nature of the causal processes has been abstracted away. One neuron sending an electro-chemical impulse to another was first likened to one transistor sending an electrical pulse to another, and now to one tank of water sending an hydraulic pulse to another. Olympia, however, has not yet graced the ballroom, for she is structurally equivalent to the brain at none of these levels.

Olympia only becomes relevant at the next level of abstraction. From the generic two events connected by a causal process we now abstract off the requirement that any causal chain connect the two and posit only a counterfactual or information-theoretic connection. Now, no physical pulse of any description need pass between two objects for the relevant relation to hold. Our chains, lying slack and inert, support no physical activity at all, but they are nonetheless transmitting information. The intact chain transmits information to the blocked machine about the state of the tape, since, had the tape been different, the block would have been pulled. The rusted chain, lying equally slack and also with no causal process coursing through it, fails to transmit that information, since it would have broken if pulled. *Information transmission between two points does not require any physical activity or causal process connecting them.*[11] So two physical systems engaged in precisely the same physical activity can be processing information differently. And now Olympia can make her entrance.[12]

[11] Cf. Fred Dretske's *Knowledge and the Flow of Information* (Cambridge: MIT, 1981), pp. 26 ff.

[12] It may be useful, now that the argument is complete, to contrast this line of investigation with some other similar arguments. As has already been noted, I do not accept Searle's Chinese room case, since it does not establish that no mental state supervenes on the activity of the man in the room, only that any such state would be disjoint from that of the man in the room. The conclusion at which Olympia's tale arrives, however, is exactly that which Searle set out to prove: having a mind is not merely a matter of instantiating a program.

Arnold Zuboff, in "The Story of a Brain" [*The Mind's I*, D. Hofstadter and D. Dennett, eds. (New York: Basic Books, 1981), pp. 202–212], subjects a cerebrum to radical disintegration in order to undermine the notion that "what decisively controlled any particular experience of a man—controlled whether it existed and what it was like—was the state of his nervous system . . ." (p. 202). Zuboff first suggests that some of the electrical connections within the brain might be stretched, resulting in a system more spread out but still capable of the same patterns of neural activity. Then, in order to overcome time-lag problems that the stretching would introduce, the direct causal connections between neurons are cut completely, being

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The doleful tale of Olympia, then, does not purport to establish that things with minds must be wet and squishy. The silicon brain and the hydraulic brain may, for all we have said, be conscious. Nothing has been said for the supervenience theses associated with these levels of abstraction, but neither has any objection been brought against them. Olympia does show, however, that consciousness cannot supervene on the computational activity of the brain. Let us turn now to a careful consideration of exactly what else Olympia shows.

V. ASSESSING THE DAMAGE

Throughout our investigations so far, we have focused on necessary and sufficient conditions for consciousness. Consciousness is intrinsically fascinating, but it is also particularly well-suited to this line of inquiry, because it is an occurrent mental phenomenon. Toothaches and other subjective experiences are events—they take place and are completely realized at or through a particular period of time. This

---

replaced by "impulse cartridges" that stimulate the neurons at the same time that they would have been stimulated if the causal connection held. The collection of neurons is still supposed to display the same "pattern of activity" in this state. Finally, the individual causally isolated neurons are scattered across the universe. The claim that the same patterns of neural firing still support any experience is shown to have unacceptable consequences.

Zuboff's case poses problems only for a particular interpretation of what constitutes a "pattern of neural activity." This is not an interpretation that the computationalist would accept. Once the causal connections between neurons have been cut, the collection of neurons (even with the "impulse cartridges") would no longer instantiate the same program. For the program, as we have seen, specifies subjunctive connections between machine states, and hence between the elements that constitute the system. But these subjunctive connections are determined by the physical and causal links between the neurons. Originally, counterfactuals such as "Had neuron *A* not fired, neuron *B* would not have fired (although it did)" are made true by the direct causal link between *A* and *B*. Once the cartridges are introduced, *B* will fire independently of what *A* does, so the counterfactual will fail.

Indeed, for any functionalist view, the relevant identity of any given neuron in the system is determined by its functional–and hence causal–role in the system, not by some tag or name attached to the particular neuron. If I switch two neurons in your brain, then exhibiting the same pattern of neural activity is not having the same particular individual neurons fire in the old sequence. Rather, you exhibit the same pattern of activity only if each switched neuron fires when its counterpart used to fire. Each inherits the role of the counterpart in defining the pattern when it inherits the functional role of the counterpart. But, in Zuboff's disintegrated brain, there are no functional roles at all, and hence no "patterns of activity."

Far from causing the computationalist problems, Zuboff's case might be used to support a computational approach. Zuboff's notion of "pattern of neural activity" is clearly too weak to capture the important aspects of brain function. Computational structure provides a stronger criterion, one which defeats Zuboff's gambit, since computational structure is not preserved through the disintegration process. Olympia, however, shows that even this stronger criterion of identifying "patterns of neural activity" is too weak. We must take into account not only the subjunctive connections between parts of the brain but the active causal connections as well, for the "patterns" defined solely by program structure cannot guarantee experience.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

fact strongly supports the thesis that they supervene on the physical events and processes that also occur through that time.

Other important mental properties, such as intelligence and understanding and intensionality, seem much more dispositional than consciousness does.[13] It is plausible to suppose that an individual's present intelligence depends on counterfactuals in a way that his present tickles and itches do not. The more dispositional a property appears, the easier it is to contend that it supervenes not only on the activities but also on the dispositions of a system. And, since the dispositions of Olympia are changed by the presence or absence of the second block, Olympia may not threaten a computational account of these other properties. The computationalist may then drop consciousness from his list of potential quarry—such a spooky phenomenon is usually not high on the list anyway—and be content to occupy himself with accounts of intelligence, understanding, intensionality, and the like. Perhaps his computers will never experience anything, but still someday they will think and speak with meaning. These goals seem quite magnificent enough.

I do not think that the outlook for the computationalist is quite so rosy. Although intelligence is more a capacity than an occurrent phenomenon, still capacities are realized in particular acts and activities. So, if having a certain program is sufficient for being intelligent, then solving a particular problem by operating according to the program is sufficient for solving the problem intelligently, and, presumably, running some nontrivial program is necessary for solving it intelligently. So let us build a version of Olympia in which the movement of the armature will instantiate the bit of that program used in solving that particular problem. (Recall that, for any given program and any finite sequence of operating that program on a given input, we can build Olympia to have that program and to operate that sequence just by the motion of the armature.) When the armature moves with the secondary blocks not in place, Olympia solves the problem intelligently, but, if we move the blocks over a few centimeters and repeat the process, she will solve the problem stupidly. Or let the output of the program be interpreted as linguistic behavior, the program designed to impart linguistic understanding. Blocks

[13] There is a sense, I think quite legitimate, in which intelligence, understanding, and intensionality all presuppose consciousness, and hence according to which the failure of computational theories of consciousness would automatically rule out computational accounts of any of these. Since most of the artificial intelligence community maintains a studious silence on questions of consciousness, I infer that they do not use these terms in this sense. The following discussion is directed toward this other sense, the sense, for example, of "belief about the ambient temperature" according to which a thermometer can have such a belief. I do not pretend to understand this other sense.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

out, and Olympia speaks with true intensionality and insight; blocks in, and she produces the very same strings in the very same way but does not mean anything by it.[14]

Ultimately, Olympia has demonstrated that computationalism, which first seemed to be a version of functionalism, actually passes to a level of abstraction which undercuts functionalism's central insight. The functionalist demands that we look beyond the surface behavior to the causal structure of the processes that produce it. A system must not only get the output right; the causal structure responsible for the output must also be of the right form. When Olympia runs π on τ, whether in writing a sentence or solving a problem, all of the blocked machinery, including the presence or absence of the secondary block, plays absolutely no causal role in producing that behavior. So, for the functionalist, the presence or absence of the block, or of the machinery as a whole, can make no difference to the intelligence, intensionality, or consciousness of the system during that computation. But, for a computationalist, the detailed structure of these inert and silent jungles of machinery makes all the difference. The verdict of the causal role functionalist is immensely the more plausible.

In *The Sandman,* when Nathanael discovers Olympia's true nature, he goes mad, finally casting himself down from a tower in a fit of violent rage.[15] We need not counsel the computationalist to anything so drastic. Our Olympia demonstrates that running a particular program cannot be a sufficient condition for having any form of mentality. But computational structure may yet be a necessary condition for consciousness or intelligence. Certainly, nothing in these arguments suggests that consideration of counterfactuals or dispositions is irrelevant to questions of mentality and consciousness. And one way, albeit not the only way, of specifying some of the dispositions of an object is by providing a description of its computational structure. If a system must be describable as running a certain kind

[14] This form of argument should sound familiar to the AI enthusiast, for it is an echo of the very arguments AI proponents use. Witness Pylyshyn's deployment of the silicon brain argument which concludes: "Thus if more and more of the cells in your brain were to be replaced by integrated circuit chips, programmed in such a way as to keep the input-output *function* of each unit identical to that of the unit being replaced, you would in all likelihood just keep right on speaking exactly as you are doing now except that [according to Searle] you would eventually stop *meaning* anything by it" (Searle, *op. cit.*, p. 442). I leave the reader to consult her or his own judgment about the strength of the intuition that replacing protoplasm with silicon cannot rob one of intensionality compared with the intuition that putting the secondary blocks in cannot rob Olympia of hers.

[15] Actually, Nathanael's final insanity is brought on when he discovers something (but what?) about the nature of Klara. I leave you, gentle reader, to consider the implications of this remarkable fact.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

of program in order to be intelligent, however, it is a further constraint that the counterfactual relations that constitute the program structure must be supported directly by the nature of physical activity in the system. Computational and causal structure are intimately intertwined in brain activity, a fact that any acceptable theory of mind must reflect.

The segregation of causally-active from counterfactual–supporting structure in Olympia is bought at a great price. In order to minimize the activity involved in running π on τ, Olympia had to be bloated with a vast galaxy of apparently redundant machinery. Indeed, one way to understand Olympia's structure is by analogy: Olympia creates the illusion of being Klara in action in much the same way that a movie creates the illusion of motion on the screen. Like still frames of celluloid, each blocked copy of Klara has captured in an inert object a moment in the career of her operation. And as the projector's light brings momentarily to life each still photo in sequence, so the movement of the armature, the successive unlatching of the floats, momentarily makes the whole dispositional structure of each copy of Klara the dispositional structure of Olympia herself. The problem with computationalism is that it does not contain the conceptual resources to distinguish the flickering illusion from reality.

Only a philosopher would consider instantiating a program in such a perverse way as in Olympia. Perhaps the pragmatic constraints that govern real AI research will naturally drive workers to create instantiated programs with just the right blending of computational and causal structure to support consciousness. But as yet we have no proof that pragmatic constraints should have any tendency to lead to this fortuitous result. At present we know neither in what direction the demands of working with digital computers exert pressure nor in what direction true artificial intelligence lies.

We must consider carefully whether the computational level of abstraction plays any part in defining consciousness and intelligence, and, if it does, what other level of description must also be invoked. Perhaps we must speak of electro-chemical structure, perhaps only more generically of the patterns of causal process in a system. Olympia has shown us at least that some other level beside the computational must be sought. But, until we have found that level and until we have explicated the relationship between it and computational structure, the belief that pursuit of the pure computationalist program will ever lead to the creation of artificial minds, or to the understanding of natural ones, remains only a pious hope.

TIM MAUDLIN
Rutgers University

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

BOOK REVIEWS

Philosophical Papers, volume II. DAVID LEWIS. New York: Oxford University Press, 1987. xvii, 366 p. Cloth $34.50, paper $18.95.

Sportswriters are fond of reminding us that a team of champions is not necessarily a champion team. But this volume of essays is both a collection of important papers and an important collection. The individual papers display the characteristic Lewis virtues: clarity, penetration, ingenuity, and technical expertise displayed where and only where it needs to be. Also, an enviably high proportion of the papers contain agenda-setting contributions to major philosophical problems. The collection, particularly when taken together with volume I of his *Philosophical Papers,* constitutes an impressively comprehensive articulation of a metaphysic which might well be called *Humean physicalism*—at least as far as it concerns *our* world. The metaphysic is spelled out in the substantial introduction to volume II. Also, volume II, like volume I, contains very helpful, and in some cases quite long, postscripts to a good number of the essays. These postscripts contain clarifications, second (or later) thoughts, replies to critics, and the like.

The papers in volume II are grouped under the four headings: "Counterfactuals and Time"; "Probability"; "Causation"; and "Dependence and Decision." I am not going to attempt to summarize them. Most of them will be familiar to many readers of this JOURNAL (possible exceptions are "Events" and "Causal Explanation" which make their first appearance in volume II). I am going to take up an issue which Lewis brings up in "A Subjectivist's Guide to Objective Chance" (83-114), and to which he returns in the introduction. He sees the issue as raising a major problem for what he calls Humean supervenience. I agree that there is a problem, but it seems to me to be not so much one for Humean supervenience as one for our understanding of chance in physics.

First we need some stage setting, and in providing it I shall simplify in the belief that the simplifications do not affect the substantive point and help prevent the details getting in the way of the right perspective on the problem (for the details, see a "Subjectivist's Guide").

Lewis believes in objective chance, understood as distinct from either credence or long-run frequency, and as making sense in the single case. He does so because of the role of chance in quantum mechanics (though he grants that this is very much a defeasible

0022-362X/89/8608/433–437 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

reason for believing in chance given the foundational problems in quantum theory). Chance on his conception is time- and history-dependent. Consider the familiar example of a coin toss, say, at noon today. The chance of heads at noon will in general vary with time as we approach noon, depending on what happens. Perhaps at 9 a.m. the chance of heads at noon is fifty percent, but at 10 a.m. the project of tossing the coin at noon is abandoned and the chance of its landing heads at noon drops sharply (as, indeed, does its chance of landing tails). Then perhaps Sly Pete decides at 11 a.m. to toss the coin, and from then until noon the chance of heads is eighty percent. But, once the coin is tossed at noon and either lands heads or lands tails, the chance of heads is then and forever 1 or 0—1 if it in fact landed heads, and 0 if it in fact landed tails. This dependence in the way chance evolves over time on what happens can be recorded in what Lewis calls history to chance conditionals—an example would be one saying that, should Sly Pete decide at 11 a.m. to toss the coin at noon, the chance from 11 a.m. until noon of heads at noon would be eighty percent. These history to chance conditionals are not, of course, material conditionals. They must be whatever sort of conditional records dependencies. Finally, on Lewis's conception, if determinism is true, all chances are 0 or 1. (But Lewis does not accept the argument that, as the coin is effectively a deterministic system, this means that its chance of landing heads must be 0 or 1. He sees no particular plausibility in the claim that, although micro-physics is most likely indeterministic, coin tosses and the like are effectively determined in outcome.)

Lewis holds that there is one thing we know above all else about chance: that beliefs about chances of outcomes strongly constrain degrees of belief or credences. He refers to this as the *principal principle*. I shall illustrate the leading idea in terms of our example. If I am certain that the chance of heads at noon is fifty percent and that is all I know, then fifty percent is the only credence to have in the coin's landing heads. What is more, even if I know a lot more than that the chance of heads is fifty percent, fifty percent will still very often be the right credence to have. Perhaps I know that the coin has been tossed many times in the past and has landed heads ninety percent of the time. This no doubt casts doubt on the chance of heads being fifty percent, but, if I am nevertheless still sure that the chance is fifty percent, then that is the credence to have. That is, if we use C for credence and P for chance,

C[Heads at noon/P(Heads at noon = fifty percent) & . . .] = fifty percent

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

holds for a large class of substitutends for the dots. Lewis calls admissible evidence anything that can be substituted for the dots without destroying the identity of credence with chance, so what we have just said can be put by saying that a lot of evidence is admissible. In particular, Lewis holds that any evidence about what has happened up to the time in question, t, say, is admissible, that is, history to t is admissible; and so also is evidence about the dependence of chance on history, the history to chance conditionals we referred to earlier, provided of course that the conjunction of the history together with the history to chance conditionals and the posited chance is consistent. (What is under the slash must be consistent.) Speaking generally, the leading idea is that the credence conditional on any consistent conjunction of a claim about chance with one or both of claims about history to the time in question and history to chance conditionals is identical with the claimed chance.

We are now in a position to address the issue that concerns Lewis. He wants to be able to accept a Humean picture of the world—"all there is to the world is a vast mosaic of local matters of particular fact, just one little thing and then another. . . . For short, we have an arrangement of qualities. And that is all. . . . All else supervenes on that" (ix-x, but see the whole of the Introduction for refinements and qualifications). His concern is how to accommodate chance into this Humean picture. As he notes, "the only hope . . . is that chances might somehow supervene on the arrangement of qualities" (xiv). This is surely right. If chance is to fit inside the Humean picture, at the least it must supervene on entirely Humean ingredients.

There are two possibilities for how it might supervene. The chance at t might supervene on matters of particular fact up to t. For instance, the chance just before noon of heads at noon might supervene on matters of particular fact up to noon, that is, on history to noon. Alternatively, what happens after noon might also in general matter, so that the chance at noon supervenes on a stretch of matters of particular fact which extends past noon. The important difference between the two possibilities is that, on the first, chances at t supervene on matters of particular fact which at t have no chance of not obtaining. For, as we noted earlier, on Lewis's scheme the chance at t of something that (actually) happens before t is 1. On the second possibility, by contrast, chances supervene on matters that have at the time in question some chance of not obtaining. Lewis observes that it is hard to reconcile the second possibility with the principal principle. He develops the argument that dependence of current chance on chancy future contradicts the principal principle in a

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

number of places in a number of different guises. Here is the simplest version.

> Let T be a full specification of history up to the present and of present chances; and suppose for *reductio* that F is a nonactual future, with some positive present chance of coming about, that would give a different present distribution of chances. What is a reasonable credence for F conditionally on T? Zero, because F contradicts T. But not zero, by the Principal Principle, because it should equal the positive chance of F according to T. This completes the *reductio* (130).

Quite apart from this difficulty for the second possibility, there is the more general, and amorphous difficulty that, when thinking about chance's role in the world, we think in terms of the outcomes of chance processes. Radioactive decay is the outcome of a chance process. How can something be the *outcome* of what is in part still to come?

What about the first possibility where chance at a time supervenes on matters of particular fact up to the time in question? If this is the case, knowledge of what has happened up to the time in question by way of particular fact completely settles the credence one ought to have at that time about many happenings in the future. For on the first possibility what has happened up to that time settles without remainder all the chances that obtain at that time, and so knowledge of what has happened gives knowledge of all the chances, and so, by the principal principle, settles the credences concerning the happenings to come which have those chances. The trouble here, as Lewis observes, is to understand how knowing what has happened to date, particularly if it is a very early to date, could be so definitive about belief about the future. How could knowledge of a perhaps relatively short period of matters of particular fact settle so definitively so much about one's credence function?

I think that Lewis is right that there is a serious problem here, but is it really a problem for Humean supervenience per se? For what exactly is wrong with the just-noted consequence of the first possibility? Lewis represents it as the problem of how knowledge of an initial segment of matters of particular fact could be so restrictive concerning belief about what is to come. But what work is the word 'particular' doing here? Is it any easier to accept that knowledge of an initial segment of matters of *fact* could be so restrictive concerning belief about what is to come? (And indeed Lewis's remarks on xv suggest that he would agree that this question is no easier to answer than the first.) Anyone who is not an inductive skeptic grants that knowledge of matters of fact constrains belief about the future; what

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

it is hard to believe is that knowledge of matters of fact constrains belief about the future to the extent of ruling that there is exactly one, highly specific degree of belief to have concerning much of what is to come. But exactly this is what follows from the combination of a realistic construal of the chances spoken of in physics together with the principal principle. A realistic construal of the chances of physics makes them out to be fully determinate, objective properties of the world around us which are possessed at particular times. The principal principle tells us that knowledge of chances fixes precisely credences about much of what is to come. Ergo, knowledge of objective properties of the world at a time can force a fully specific degree of belief about the world after that time. It follows that the (or at least *a*) worrying result can be got without mentioning Humean supervenience. (Although the problem here is not one for Humean supervenience, if I am right, it is nevertheless a Humean problem. Hume famously asked how one might get an 'ought' from an 'is'; we are asking how one might get an 'ought to believe about the future' of such precision from an 'is'.) But, if the difficulty is the result of combining realism about single case chance with the principal principle, then, in view of the appeal of Lewis's claim that the principal principle embodies "all we know about chance" (86), realism about single-case chance becomes correspondingly less appealing.

FRANK JACKSON

Australian National University

Whose Justice? Which Rationality? ALASDAIR MACINTYRE. Notre Dame: University of Notre Dame Press, 1988. xi, 410 p. Cloth $22.95.

Alasdair MacIntyre's most recent book is more than just a sequel to *After Virtue,* the neo-Aristotelian attack upon modern ethics which he published in 1981. It announces some significant changes in doctrine. MacIntyre has shifted his ultimate philosophical allegiance from Aristotle to Aquinas. At the same time, he has taken back the one objection which he made against Aristotle's ethics (and which turned Sophocles, in fact, into the real hero of that book): the perception of tragic moral conflicts no longer counts as a sign of moral maturity, but as a proof of moral ignorance instead (62, 186). The most important development in this new book comes, however, at the deeper level of moral epistemology. MacIntyre's principal claim

0022-362X/89/8608/437–442 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

is that rationality—in ethics as elsewhere—is possible only within a tradition. Reasons for belief and action, he argues (6, 346, 393), cannot properly be thought to be recognizable by any rational being as such, since the justification of belief and action must always rely upon given commitments that are part of some particular tradition. The effort to make rationality consist in a critical distance toward all tradition must end, if consistent, in the loss of the means for any significant form of rational thought. Although *After Virtue* contained some allusions to this sort of epistemology, it now stands firmly at the center of his thought. Besides elaborating it in some detail, he puts this view of reason and tradition to two uses. In a series of case studies, he shows how the ideals of social justice worked out in three ethical traditions—the Platonic-Aristotelian, the Augustinian, and the Scottish—hang together with conceptions of practical reason which are characteristic of these traditions. He also charges that the apparently interminable controversies about justice in modern society follow from the absence of any coherent tradition fostering a shared conception of practical reason. Our disorientation, according to MacIntyre, springs precisely from the Enlightenment's misguided attempt to cast reason above the contingencies of tradition (6, 335, 346).

Although always informative and vigorously written, this book is of uneven value. MacIntyre's epistemology of reason and tradition seems to me to move on the right track. His historical discussion of the links between ideals of justice and conceptions of practical reason is learned and often quite insightful. But his critique of modern ethics ought to carry little conviction. In large part this is because he is unfair, assailing modern thought for commitments which he condones or conceals in his favored premodern traditions. I shall focus on the first and third of these topics, since they involve his deepest philosophical beliefs.

It is best to begin with his claim that rationality is possible only within a tradition. This view is supposed to apply to every domain of thought, though we should remember that his chief concern lies with its bearing on moral theory. Rational thought, according to MacIntyre, always begins from contingently given beliefs, beliefs which are ours simply because we belong to some particular tradition of thought (354, 360). Rational changes of belief, he continues, depend on the conceptual resources available in the tradition, both for the articulation of the problems to be solved and for the standards of what may count as acceptable solutions (167). Without offering anything like a definition of the master term 'tradition', MacIntyre does make two important points. A tradition of inquiry is "progressive,"

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

he holds (79f.), if its members can build on previous findings, can explain how earlier disagreement was possible, and can give an increasingly adequate account of the goal of the inquiry, which is an ultimate explanation of its subject matter. He also insists that traditions differ too much in their beliefs, standards, and goals for us to expect that reasons for belief developed in one tradition can be cogent to any rational being as such. Here, MacIntyre is not making the self-defeating claim that there is no idea of rationality shared by all traditions. The norm of avoiding logical inconsistency belongs to neutral ground, as does presumably the norm of progressiveness. His point is rather that this common notion of "rationality as such" is too weak to ensure that, on any substantive matter, reasons for belief as understood by one tradition will be universally acknowledgeable (166, 333, 351, 367). He does not believe that disagreements between traditions must therefore be rationally unresolvable. But we can have reason to switch our allegiance to a rival tradition, only if our own tradition has ceased to be progressive, if the rival can explain both what our tradition cannot explain and why it cannot do so, and if this explanation is cogent by the standards of our own tradition (166/7, 364/5).

This contextualist epistemology is one we should accept, once we give up (as I believe we must) the view that there is an idea of reason that is both universal and sufficient by itself to settle anything of substance. In its essentials, this is the position of pragmatism. But we would be foolish to expect from MacIntyre any kind words about modern thinkers. The error of pragmatism, he charges, is that it aligns the concept of truth with "an idealization of the concept of warranted assertibility" (169), and not with "something external" to one's "own particular scheme of concepts and beliefs" (170), as in Aquinas's correspondence theory of truth. This is an odd distinction for MacIntyre to draw, though his wish to invoke it is symptomatic of a paradox in his attitude toward tradition. Connecting truth with idealized warranted assertibility seems to combine two advantages, which he, too, should value: (a) we can grant that beliefs which currently count as justified may be ideally found to be false, and in this way truth is something external to our present beliefs; but (b) we can attach a real sense to their possible falsity, one which makes a difference to the course of our thinking, since it would consist in these beliefs failing to square with those of our present standards which we project as ideal norms of justification. What more external notion of truth could MacIntyre want, given his principle that rationality is possible only within a tradition, and given his account of when we have reason to abandon a tradition?

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Aquinas, of course, did affirm a more transcendent notion of truth. But this is precisely because he did not share the tradition-bound, contextual view of reason which MacIntyre advocates. For him, as indeed for MacIntyre's other heroes—Plato, Aristotle, and Augustine—the task of reason was to rise above historical contingencies, and the aim of moral philosophy was to discern some fixed set of norms which any rational agent as such must acknowledge (see *Summa theologiae* I–II, Q. 94, a. 4). At times (8, 252), MacIntyre concedes this fundamental point of difference between his own epistemology and the moral traditions he champions. But he does not seem to appreciate that it has two important consequences. First, it can scarcely be fair to single out the Enlightenment for blame, as he does, just because it sought to pry reason loose from tradition. In this ambition, it was at one with earlier thought, whatever other significant differences there were. Second, his epistemology precludes being a full-fledged member of one of these earlier moral traditions, at least in a form that matches their basic self-understanding. That this is nonetheless what he wishes to be is evident from his profession that he is an "Augustinian Christian" (10), as well as from his appeal to Aquinas's theory of truth. But these traditions of thought were not themselves tradition-minded, as MacIntyre is. Here is a paradox of traditionalism, which earlier defenders of tradition (Burke and de Maistre, for example) have also chosen to ignore. I do not mean to suggest that MacIntyre's epistemology is inconsistent. But he cannot consistently think of it as enabling him to be in any simple sort of continuity with earlier (pre-Enlightenment) moral theory. He must view these moral traditions as being historically conditioned in a way they did not understand themselves to be. His outlook is not a return to the past, but an innovation, and a distinctively modern one at that.

There is another important inadequacy in MacIntyre's understanding of moral tradition. It, too, leads him to be unfair in his critique of the Enlightenment and blinds him to how much of a modern he is himself. Generally, he assumes that being part of a moral tradition rules out role distance. Because reasons for belief and action cannot be acknowledged by rational beings as such (apart from the standards of some particular tradition), he believes that they should address bearers of this or that social role (141, 321, 339). The tendency of modern thought to portray the moral agent as unidentified with any concrete role is for him one more sign of its lack of a sustaining tradition. Contrary to MacIntyre, however, role immersion is not the only alternative to trying to reason in a social vacuum. A form of social life may provide its members with princi-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ples with which, by standing back, they are to measure the moral force of the demands made by social roles, and such has indeed been one of the dominant strands in modern society. Moreover, this morality of role distance has one of its most important sources in Christian thought—a fact for which MacIntyre himself supplies significant evidence, though apparently that is not his intention.

Take, for example, his interesting explanation of Aristotle's claim that the conclusion of a practical syllogism is, not a judgment, but an action. Because Aristotle saw the rational agent as "inhabiting a structured role" (141), he did not share the "characteristically modern" (140) view that the agent must make a further decision to act on the reasons he affirms in the premises of the syllogism. But what in this discussion MacIntyre brands as a modern pathology he later describes (154–156), quite correctly, as Augustine's anti-Aristotelian theory that the rational agent must not just perceive the good, but also *will* to pursue it, if he is to act to achieve it. This is not the only way in which Christian thought encouraged the role distance MacIntyre finds so shallow a perspective in modern ethics. The "modern liberal view," he claims (146), unlike the Aristotelian, intends that the norms of justice apply to "the relationships of human beings as such." But did not Christianity also aim its message at a universal audience? In his discussion of Augustine, he says just that, though without retracting his earlier indictment of modernity: "the law of the *civitas Dei* is by contrast [with Aristotle] a law for all mankind" (163). Part of MacIntyre's trouble here is that once again he has unfairly criticized modern thought for what he approves in an earlier tradition. Another part is that he does not distinguish three things that the idea of a universalist morality may mean: (1) a morality of duties binding on all rational agents as such; (2) a morality of duties toward all rational agents as such; and (3) a morality that all rational agents as such will acknowledge as binding on them. MacIntyre's epistemology excludes (3), but not (1) and (2). A moral tradition may be historically conditioned (contrary perhaps to its own ambition) and yet transcend role morality. This has been a central continuity between Christian and modern ethics. In professing to be an Augustinian Christian, MacIntyre is making himself the modern he claims not to be.

A frustrating, but also endearing feature of MacIntyre's book is that the mistakes he appears to make in some passages he corrects in others, though without admitting that he is doing so. I have already noted two of these reversals. Here is another: If pre-Enlightenment moral traditions turn out, on his own account, to contain two elements that he makes responsible for the Enlightenment's lack of

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tradition—the ambition to raise reason above tradition and the construction of a role-independent morality—we may well ask why modern ethics should not equally qualify as a moral tradition. In fact, MacIntyre concedes that it is one (10, 335, 345/6). At this point, however, he is left with only one way of unfavorably contrasting modern ethics with earlier traditions. Modern thought is rife, he asserts, with interminable moral controversies (5, 335, 344, 349). This may be true, but earlier traditions had their own controversies, which often were terminated only to the extent that these traditions themselves came to an end. MacIntyre says nothing about the persistent Stoic critique of Aristotelian ethics, nor about the fact that Aquinas's synthesis of the Aristotelian and Augustinian moral traditions no sooner appeared than it met with official condemnation in 1277 (because of its alleged neglect of the will) and became an object of attack for Scotus and Ockham. The history of moral thought is one of ceaseless controversy, in which great minds are divided on fundamental issues. Of course, some of the disputed theories may nonetheless be correct. But this is true of modern theories as well.

Although MacIntyre's critique of modern ethics in the name of earlier traditions is at all points unfair, a real merit of his book is that it provides evidence to disarm this critique. In the end, his historical honesty wins out, as well as his own buried commitment to modernity. Still, MacIntyre should conquer his impulse to regard the history of modern ethics as simply the history of error. This approach ignores important continuities between modern and earlier thought. It also misses the effort of modern thinkers to remedy the real defects in earlier traditions. Only when our working hypothesis is that modern ethics is a mix of error and insight are we likely to get a fair view of its worth.

CHARLES LARMORE

Columbia University

Meaning and Method in the Social Sciences. PAUL A. ROTH. New York: Cornell University Press, 1987. 250 p. Cloth $26.95.

The fact of meaning has been thought by many to pose special problems for the project of constructing scientific social theories. Paul Roth claims that, once one reveals the falsity of the central assump-

0022-362X/89/8608/442–446 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tions underlying this thought, then the problem posed by meaning can be diagnosed as a pseudo problem and all will be happy. The two major (false) assumptions are (a) that there is a fact of the matter about meaning, and (b) that there is a unity of method in science. The exact relation between these two assumptions is obscure, so I shall consider them in turn.

Roth claims that those who have opposed the project of a social science have done so on the grounds that explanations that invoke meanings are different from the predictive type of explanation to be found in natural science. These antiscientists are thus committed to meaning realism, the doctrine that "some intentional or subjective set of beliefs (and not social and behavioral conditions) determines the significance or meaning of an utterance" (2). W. V. O. Quine's argument for indeterminacy of translation proves, however, that meaning realism is false, so the antiscientists are defeated. [The contingent connection here between assumptions (a) and (b) is that those who assumed meaning realism also believed that there was a single method to be pursued in the study of society, a method distinct from that used in natural science.] Clearly the argument for indeterminacy is crucial, and Roth properly takes some time defending his own construal of it. On route he also claims that the argument leading to indeterminacy provides compelling reasons for denying assumption (b) and for adopting methodological pluralism. The central premise in this argument is that theoretical sentences have their meaning and their evidence only in virtue of their place within a theory (Quine's holism). Roth's construal of the argument for indeterminacy is unusual insofar as he takes this holism to be almost sufficient for the conclusion of indeterminacy. It would be sufficient except that there has to be an explanation as to why indeterminacy holds for semantics and not for physics. The explanation proffered is that language learning presupposes that there be "intersubjectively available stimuli to which speakers or potential speakers of a language are capable of responding alike" (19). From inside our web of belief, we discern the necessity for shared stimuli; such necessity provides us with a warrant for a belief in objective evidence, and positing a fact of the matter explains why such evidence is intersubjectively available. This strategy, it is alleged, does not work for semantics, as no similar case can be made for the necessity of some special intentional evidence to which semantic theories must be beholden, so indeterminacy results.

Roth claims that this route to indeterminacy avoids the problem besetting those which use underdetermination as a central premise, that of distinguishing the case of semantics from that of physics. As both are underdetermined, it appears to be plain prejudice against

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

intentionality to deny a fact of the matter to semantics but allow it to physics. Roth's denial of the necessity of intentional evidence in the semantic case can be reversed, however; if such a denial leads to indeterminacy, then obviously intentional evidence *is* necessary. Roth aims to prove non-necessity by showing that it is pointless to require of any semantic theory that it deliver translations more determinate than what is constrained by the evidence provided by our dispositions to respond. It is clearly crucial that these dispositions be nonmentally characterized, otherwise the thesis of the non-necessity of intentional evidence will be disproven. Roth does not provide much by way of argument which will dispel doubts that the behavioral level of description will provide anything like what is required of evidence for any half-way plausible semantic theory. Indeed, it would appear that Roth is committed to some intentional descriptions in the characterization of the shared stimuli which form both the basis of our communicative abilities and the objective evidence whereby we can be assured that physics has a fact of the matter. Roth declines to follow Quine in a descent to neurophysiological description, claiming that this would not be able to account for the social nature of language, and so Quine would be committed to a form of linguistic solipsism. Roth prefers to use perceptual similarity as the necessary basis, but glosses this as involving behavioral similarity. If this similarity at the behavioral level does not connect with any of the beliefs of speakers (something which perceptual similarity, ordinarily understood, would imply), it would also appear to result in linguistic solipsism.

Even if we were to grant the case against meaning realism, there would still be problems with the rest of Roth's case for methodological pluralism. For one thing, it is unclear exactly what the theoretical connection between meaning realism and methodological unity is supposed to be. Those accused of succumbing to the wiles of meaning realism are Peter Winch and those of his opponents who believe in the "epistemological unity of mankind" (e.g., Martin Hollis and Steven Lukes). The problem that divides them is whether meaning is local or universal; if there is no fact of the matter about meaning, then the problem is supposed to disappear. Let us suppose that meaning realism is false. The problem dividing Winch from his opponents could then be put thus: in the different translations of the alien, must there be some element, such as general conformity to the law of noncontradiction, common to all those translations and to us? This way of putting the dispute does not depend upon any belief in meaning realism, which suggests that the dispute will survive the loss of that belief. In addition, it cannot be the case that a belief in there

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

being a fact of the matter about meaning is essentially connected to any thesis about methodological unity. Roth himself believes both that there is a fact of the matter in natural science and that methodological pluralism should nevertheless prevail. It seems that, as far as any thesis about methodological pluralism is concerned, there being a fact of the matter about meaning or not drops out as irrelevant.

The discussion of methodological unity versus pluralism is bedevilled by a lack of clarity concerning the relevant notion of method. The advocates of methodological unity usually distinguished methods from techniques, and meant by 'method' nothing more detailed than suitable responsiveness to evidence provided by observation and experiment. To be sure, there were debates over what 'suitable' meant: the positivists (especially Rudolf Carnap) generally intended it to mean establishing hypotheses with a high degree of confirmation, whereas Karl Popper preferred a high degree of corroboration. If one put the unity thesis as "scientists must make acceptance and rejection of hypotheses sensitive to predictive success," then it would appear that Roth is a unifier. Although he suggests that we could "change our whole notion of rational justification," this change is subject to its serving the aim of science, which is successful prediction. In this part of the book (ch. 3), Roth seems to go along with this view of science, the discussion focusing on the differences between Quine and Paul Feyerabend on how best to achieve successful prediction, Quine's conservatism contrasting with Feyerabend's plea for theoretical proliferation. But these are different means toward the common end of true prediction, and there is presumably a right answer as to which affords the better means toward that end.

Roth does suggest that he is committed to a more radical position: he wishes to talk of alternative rationalities, and he identifies intellectual inquiry with self-regarding actions. On the first point, he concurs with an interpretation of Socrates's method in the *Apology* which takes Socrates to be initiating a form of rational debate which was new to Greek society, a form which eschewed the traditional reliance on rhetoric in favor of argumentation relying solely on the persuasive power of truth. He also cites Jack Goody's work on the difference between oral and literate cultures as showing that there are different rationalities at work in the two types of society. These differences are usually associated with explanations as to why literate cultures were more successful at developing and enlarging scientific knowledge, however, and so do not support any position which might use these differences in arguing style in support of scientific methodological pluralism. [Roth himself talks of the trials of Socrates and Galileo as "dramas in the development of reason" (114),

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

thus indicating that he thinks that what was replaced was inferior to what was initiated.] On the second point, Roth sometimes speaks as though there is (one) scientific method, and that what he is arguing for is that there is more to rationality than adherence to that method: "the moral of the story is that, while scientific method might be the first word in rational inquiry, it is a mistake to take it as the final word" (97). One should be free to experiment with all methods of inquiry, providing there is no harm to others in so doing. Clearly there can be no objection to this, but again it does not support any radical thesis about methodological pluralism in science. Roth's own pluralism amounts to a plea for a "dynamic balancing of experimentation and constraint" (112), with there being no context-free way of determining when experiment or constraint is appropriate, a conclusion which most nonfoundationalists will find agreeable.

Despite the above criticisms, there are some valuable discussions of the positions of Richard Rorty and Feyerabend, as well as two chapters containing a detailed critique of the strong program in the sociology of science. The parts of this book are more interesting than the supposedly radical holism that binds them together.

GRAHAM MACDONALD

University of Bradford

NEW BOOKS: TRANSLATIONS

ADORNO, THEODOR W.: *Kierkegaard: Construction of the Aesthetic.* ROBERT HULLOT-KENTOR, trans. Minneapolis: Minnesota UP, 1989. xiii, 166 p. Cloth $35.00, paper $14.95.

DERRIDA, JACQUES: *Limited, Inc.* SAMUEL WEBER and JEFFREY MEHLMAN, trans. Evanston: Northwestern UP, 1989. viii, 160 p. Cloth $24.95, paper $10.95.

DUSSEL, ENRIQUE: *Ethics and Community.* ROBERT R. BARR, trans. Maryknoll, NY: Orbis, 1988. xii, 260 p. Cloth $26.95, paper $13.95.

HABERMAS, JÜRGEN: *On the Logic of the Social Sciences.* SHIERRY WEBER NICHOLSEN and JERRY A. STARK, trans. Cambridge: MIT, 1988. xiv, 220 p. Cloth $22.50.

HEGEL, G. W. F.: *Introduction to the Philosophy of History.* LEO RAUCH, trans. Indianapolis: Hackett, 1989. xvi, 106 p. Cloth $22.50, paper $4.95.

HUEBSCHER, ARTHUR, ed.: *Schopenhauer: Critical Debates (1809–1818).* E. F. J. PAYNE, trans. New York: Berg, 1988. xxxi, 493 p. Cloth $68.00.

HUMBOLDT, WILHELM VON: *On Language: The Diversity of Human Language Structure and Its Influence on the Mental Development of Mankind.* PETER HEATH, trans. New York: Cambridge, 1988. lxix, 296 p. Cloth $49.50, paper $14.95.

JACOBI, F. H.: *The Spinoza Conversations between Lessing and Jacobi: Text with Excerpts from the Ensuing Controversy.* GERARD VALLEE, J. B. LAWSON,

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

and C. G. CHAPPLE, trans. Lanham, MD: UP of America, 1988. vi, 174 p. Cloth $26.50, paper $13.75.

KAUTSKY, KARL: *The Materialist Conception of History.* JOHN H. KAUTSKY, ed. and trans. New Haven: Yale, 1988. lxx, 558 p. Cloth $55.00.

LYOTARD, JEAN-FRANCOIS: *The Differend: Phrases in Dispute.* GEORGES VAN DEN ABBEELE, trans. Minneapolis: Minnesota UP, 1988. xvi, 208 p. Cloth $35.00, paper $14.95.

MCHEDLOV, M.: *Religion in the World Today.* NATALIA BELSKYA and FELIX KHOLMOGORTSEV, trans. Moscow: Progressive, 1988. 236 p. Paper $4.95.

PIAGET, JEAN, and ROLANDO GARCIA: *Psychogenesis and the History of Science.* HELGA FEIDER, trans. New York: Columbia, 1989. xi, 309 p. Cloth $37.50.

SCHELLING, FRIEDRICH WILHELM JOSEPH VON: *Ideas for a Philosophy of Nature.* ERROL E. HARRIS and PETER HEATH, trans. New York: Cambridge, 1988. xxv, 294 p. Cloth $49.50, paper $15.95.

SCHLEIERMACHER, FRIEDRICH: *On Religion: Speeches to Its Cultural Despisers.* RICHARD CROUTER, trans. New York: Cambridge, 1988. xii, 231 p. Cloth $37.50, paper $12.95.

SCHWEITZER, ALBERT: *A Place For Revelation: Sermons on Reverence for Life.* DAVID LARRIMORE HOLLAND, trans. New York: Macmillan, 1989. xiv, 130 p. Cloth $18.95.

SIMPLICIUS: *On Aristotle's* Physics *6.* DAVID KONSTAN, trans. Ithaca: Cornell, 1989. 181 p. Cloth $36.50.

SYUSYUKALOV, B. I., and L. A. YAKOVLEVA, eds.: *A Handbook of Philosophy.* STEPAN APRESYAN and LUDMILA LEZHNEVA, trans. Moscow: Progress, 1988. 142 p. Paper $3.95.

VENETI, PAULI: *Logica Magna: Secunda Pars: Tractatus de Obligationibus.* E. JENNIFER ASHWORTH, trans. New York: Oxford, 1988. xvi, 409 p. Paper $98.00.

NOTES AND NEWS

Dialectic and Humanism, the Polish Philosophical Quarterly, announces the First International Symposium on Universalism, entitled "Metaphilosophy and New Thinking: The Impact of the Studies on the Meaning of History and the Meaning of Life on the Development of a Global, Peaceful World," to be held in Warsaw, November 6–9. Further information may be obtained by writing *Dialectics and Humanism*, ul. Miodowa 10, 00-251 Warsaw, Poland.

The Department of Logic and Philosophy of Science at the University of the Basque Country, Spain, will host an International Symposium on Structures in Mathematical Theories, to be held in San Sebastian, September 24–28, 1990. The symposium, which offers an open platform for mathematicians and philosophers and historians of science to discuss the singularity of mathematical theories, will consist of five sections: mathematical theories and em-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

pirical theories, applications of mathematical theories, history and sociology of mathematical theories, methods of research into mathematics, and structures of mathematical theories. Inquiries may be directed to Javier Echeverria, Dept. of Logic and Philosophy of Science, Universidad del Pais Vasco, Apartado 1249, 20080 San Sebastian, Spain.

The Social Philosophy and Policy Center of Bowling Green State University will hold a conference on Ethics, Politics, and Human Nature, on October 19–21. Conference participants will include Bernard Williams (Berkeley), David Gauthier (Pittsburgh), Jonathan Barnes (Oxford), Alexander Rosenburg (Riverside), H. Tristam Engelhardt, Jr. (Rice), Michael Ruse (Guelph), Christina Hoff Sommers (Clark), Andrew Oldenquist (Ohio State), Lester Hunt (Wisconsin), and Richard Epstein (Chicago). Further information may be obtained by writing Kory Tilgner, Social Philosophy and Policy Center, Bowling Green State Univ., Bowling Green, OH 43403-0188.

Harvard Law School will offer four or five Liberal Arts Fellowships to college and university teachers in the arts and sciences for the academic year 1990–91. Holders of these fellowships will have the title Fellow in Law and . . . , depending upon their particular disciplines. The purpose of the fellowships is to enable teachers in the social sciences or humanities to study fundamental techniques, concepts, and aims of law, and fellowship holders will take at least two first-year courses in law, in addition to more advanced courses, and will participate in a joint seminar. Applications, which should include a curriculum vita, statement of purpose, and two letters of recommendation, should be submitted by January 15, 1990, to the Chair, Committee on Liberal Arts Fellowships, Harvard Law School, Cambridge, MA 02138.

The Society for Women in Philosophy, Eastern Division, announces its Fall conference, to be held at the University of Massachusetts/Amherst on October 21–22. Speakers will include Ann Ferguson and Vicki Spelman. Papers and work in progress on all topics are invited, and graduate students are encouraged to participate. Abstracts may be submitted in duplicate by September 7 to Becky Homes, 24 Berkshire Terrace, Amherst, MA 01002.

The Stanford Humanities Center will be offering eight external fellowships for 1990–91, intended for postdoctoral scholars and teachers in the humanities. The fellowships are primarily intended to enable fellows to pursue their own research and writing; however, recipients are also expected to devote about one-sixth of their time to teaching or in some other way contributing to the intellectual life at Stanford. The deadline for application is December 1. Further information may be obtained by writing Morton Sosna, Stanford Humanities Center, Mariposa House, Stanford Univ., Stanford, CA 94305-8630.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVI, NO. 9, SEPTEMBER 1989

THE APPLICATION OF MATHEMATICS TO NATURAL SCIENCE*

A number of great physicists have expressed amazement that mathematics should be applicable to physics. Eugene Wigner[1] says, "[t]he miracle of the appropriateness of the language of mathematics for the formulation of the laws of physics is a wonderful gift which we neither understand nor deserve" (*ibid.*, p. 237). Similarly, Heinrich Hertz[2]:

> One cannot escape the feeling that these mathematical formulae have an independent existence and an intelligence of their own, that they are wiser than we are, wiser even than their discoverers, that we get more out of them than was originally put into them (*ibid.*, p. 99).

Steven Weinberg[3]:

> It is positively spooky how the physicist finds the mathematician has been there before him or her (*ibid.*).

* This paper has benefited from many critics: Daniel Amit, Paul Benacerraf, Bernard Berofsky, Sylvain Cappell, Yael Cohen, Shmuel Elizur, Bas C. Van Fraassen, David Malament, Nathan Malkin, Sidney Morgenbesser, Yuval Ne'eman, Ithamar Pitowsky, Carl Posy, Hilary Putnam, Stewart Shapiro, Barry Simon, Shlomo Sternberg, and Larry Zalcman. It was prepared with the aid of a grant from the Israel Academy of Arts and Sciences, for which I express my gratitude.

[1] "The Unreasonable Effectiveness of Mathematics in the Natural Sciences," reprinted in *Symmetries and Reflections* (Bloomington: Indiana UP, 1967).

[2] Quoted in Freeman Dyson, "Mathematics in the Physical Sciences," in *The Mathematical Sciences*, Committee on Support of Research in the Mathematical Sciences, eds. (Cambridge: MIT, 1969). Compare also Richard Feynman, *The Character of Physical Law* (Cambridge: MIT, 1967), p. 171: "When you get it right, it is obvious that it is right . . . because usually what happens is that more comes out than goes in."

[3] *Notices of the American Mathematical Society*, XXXIII, 5, 250 (October 1986), p. 725.

0022-362X/89/8609/449–480

© 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Richard Feynman[4]:

> I find it quite amazing that it is possible to predict what will happen by mathematics, which is simply following rules which really have nothing to do with the original thing (*ibid.*, p. 171).

and Johannes Kepler[5]:

> Thus God himself was too kind to remain idle, and began to play the game of signatures, signing his likeness into the world; therefore I chance to think that all nature and the graceful sky are symbolized in the art of geometry (*ibid.*, p. 99).

Finally, Roger Penrose[6] characterizes the applicability of mathematics in physics as a

> . . . profound interplay between the workings of the natural world and the laws and sensitivity of thought—an interplay which, as knowledge and understanding increase, must surely ultimately reveal a yet deeper interdependence of the one upon the other (*ibid.*, p. 84).

These sentiments find almost[7] no echo in the writings of contemporary philosophers: the problem of the applicability of mathematics is either ignored[8] or dismissed.[9]

What captures the imagination of the physicists is what they view as the mysterious correspondence between mathematical concepts and the universe. It is the purpose of this essay to say what the correspondence is, what the evidence for it is, and why it requires explanation. And, despite the "realist" flavor of some of our quoted remarks, I shall try to do all this without presupposing any strong form

[4] *The Character of Physical Law* (Cambridge: MIT, 1967).

[5] Quoted in Dyson.

[6] "The Geometry of the Universe," in *Mathematics Today: Twelve Informal Essays*, Lynn Arthur Steen, ed. (New York: Springer, 1978).

[7] Nicholas Rescher's *The Riddle of Existence* (Lanham: UP of America, 1984) is an exception.

[8] The second edition of the standard anthology, *The Philosophy of Mathematics*, Paul Benacerraf and Hilary Putnam, eds., has not a single article on the subject. Benacerraf informs me that lack of material was the reason. Recent works in the philosophy of mathematics, such as Philip Kitcher's *The Nature of Mathematical Knowledge* and Hartry Field's *Science without Numbers*, discuss the applicability of mathematics as a problem that arises from the disparity between the ostensible subject matter of mathematics and that of the physical sciences. Cf. also Stewart Shapiro's "Mathematics and Reality," *Philosophy of Science*, XXX (1984): 523–548. Whatever the merits of that problem, it is a side issue here.

[9] "It is no mystery, therefore, that pure mathematics can so often be applied. . . . It is a reasonable hypothesis that pure mathematics in general is so often applicable, because the symbolic structures it studies are all suggested by the natural structures discovered in the flux of things"—Ernest Nagel, "Impossible Numbers," in *Teleology Revisited* (New York: Columbia, 1979), p. 194.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

of "scientific realism."[10] Thus, I shall no longer speak, in what follows, of the applicability of mathematics to nature, or of a correspondence between mathematics and the universe, but only of the applicability of mathematical concepts in physical theory. Realists will make the appropriate translations. Our first task, then, is to define the applicability of mathematics in the natural sciences.

I.

Mathematical concepts, obviously, appear in the basic laws of physics, where they are used, for example, to predict the future as resembling the past. Because some of these predictions are amazingly accurate,[11] and because the concepts that make these predictions possible were often not formulated for that purpose,[12] Wigner speaks of the "unreasonable effectiveness of mathematics in the natural sciences," calling it a "miracle" (*op. cit.*).

Wigner's essay, unfortunately, has been largely ignored by philosophers, for one of three reasons:

(1) They are not persuaded by individual examples that the applicability of mathematics to physics is a general phenomenon (what about all the mathematical concepts that are not applicable, they ask).

(2) They agree that the applicability of mathematics is a phenomenon, but are not persuaded that there is anything mysterious about it. Mathematical concepts, they say, arise out of experience (Kantians say that the mathematical concepts are constitutive, in some way, of experience) and thus it is reasonable that these same concepts should be suited to describe experience. "Naturalists" add that a preference for symmetry may well be a product of natural selection.

(3) They agree that the applicability of mathematics is a mystery, but say that there is nothing special about mathematics—the applicability of any conceptual schemes is a mystery.

In my opinion, each of these objections is much weaker than its proponents believe; rather than refute them, however, I shall present my own version of Wigner's thesis—one to which the objections are, I hope, irrelevant.

[10] I doubt that I shall persuade Bas Van Fraassen, however, since he regards the very drive for explanation which motivates me here as somewhat suspect. Cf. *The Scientific Image* (New York: Oxford, 1980), pp. 23–25.

[11] Newton's law of gravity is accurate to better than one part in ten thousand.

[12] Appolonius's theory of conic sections had no application in physics until Kepler used it, fifteen hundred years later, in formulating the laws of planetary motion. Riemann developed non-Euclidean geometry to solve problems in thermodynamics; fifty years later, Einstein applied it to gravitation. Elie Cartan, in his doctoral dissertation, classified the simple Lie algebras—not realizing that he was, at the same time, classifying elementary particles. The list is very long.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Wigner's thesis is about the usefulness of mathematical concepts in framing successful physical descriptions.[13] Of course, the most useful of the mathematical descriptions in physics can be used to predict future experience, because they express in what respect the future is likely to be analogous to the past. I call such analogies, between past experience and future experience, *first-order analogies;* there are higher-order analogies as well.

My version of Wigner's thesis is not a thesis about the aptness of mathematical concepts in describing experience—or physical reality, if you prefer—but a thesis about the role of mathematical concepts in discovering those very descriptions. The thesis is this: an indispensable factor in contemporary physical discovery (I am speaking of Maxwell's equations; Einstein's general theory of relativity; quantum mechanics: the elementary version, the relativistic version, and quantum field theory; as well as many other examples) has been the use of mathematical analogies; for example, physicists attempting to discover some physical description restricted their search to descriptions with the same mathematical properties as known, successful[14] descriptions. I call such analogies *second-order mathematical analogies,* because they are based on properties of descriptions, rather than on the descriptions themselves. Thus, linearity is a second-order property of (certain) equations; a preference for linear equations based on the proven success of other linear equations would be a second-order analogy.

Occasionally, the physicist resorts to a *third-order mathematical analogy,* one based upon the properties of properties of descriptions! For example, as we shall see later, Murray Gell-Mann argued formally from the existence of a certain type of symmetry in nature to the existence of an analogous, but more general, symmetry. Although Gell-Mann did not know it at first,[15] the analogy he was trying

[13] I am using the term 'description' more loosely than in contemporary logic. For example, an equation also counts as a description. A mathematical description is simply a description that contains a mathematical concept.

[14] In order to avoid getting bogged down in philosophical controversy, I shall not attempt to characterize "successful" predictions.

[15] See "Particle Theory From S-Matrix to Quarks," in *Symmetries in Physics (1600–1980)* (Proceedings of the 1st International Meeting on the History of Scientific Ideas Held at Sant Feliu de Guixols, Catalonia, Spain, September 20–26, 1983) (Barcelona: Universitat Autonoma Servei de Publicacions, 1987), pp. 473–497. I am grateful to Yuval Ne'eman for making available to me this and many other articles on the "eightfold way" and for a number of illuminating communications (written and oral) on the subject. Although Ne'eman also discovered the eightfold way, independently of Gell-Mann, his reasoning (as outlined, e.g., in "Hadron Symmetry, Classification, and Compositeness," in the Barcelona volume,

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

to draw was one between two "Lie groups," thus a third-order analogy.

Now, in some cases, a mathematical analogy has a physical basis, because some mathematical properties are "equivalent" to physical properties.[16] For example, if an equation is *linear,* then the sum of two solutions is again a solution. Equivalently, we have two causes, each acting as if the other were not present—this is the so-called "principle of superposition," which played a central role in Galileo's study of falling bodies. We may thus say that linearity—a second-order mathematical property of equations—is "equivalent" to the applicability of the principle of superposition, a physical postulate. Again, the mathematical property of an equation known as "covariance" was regarded by Einstein as "equivalent" to "Mach's principle," according to which absolute space has no causal efficacy whatsoever.[17] Any analogy based on linearity or covariance, therefore, is physically based. On the other hand, a second-order, or third-order, mathematical analogy which is not physically based will be called a *formal analogy.* For example, as I shall argue later, Gell-Mann's analogy had no physical backing at the time, and was therefore a formal analogy.

Thus, my thesis, amplified, is: an indispensable factor in modern physical discovery in the past one hundred years has been the use of formal mathematical analogies. The bulk of this essay is simply the evidence for this thesis.

Please note the relativity to time in my version of Wigner's thesis; that is, there are a number of mathematical analogies which, though formal at the time they were made, were "physical" later on. This is the case, for example, with the analogies drawn by Gell-Mann: their success is explained by the later theory of "quarks,"[18] due, of course, also to Gell-Mann.[19] But in 1960—when Gell-Mann drew his anal-

---

pp. 501–524) does not seem to have involved mathematical analogies—and therefore is not discussed in the text.

[16] What is meant, of course, is that, on certain background assumptions, a description having a certain mathematical property must describe objects of a certain type.

[17] I shall assume that Einstein was right, merely for the sake of argument.

[18] This needs to be qualified: according to the present theory of quarks, the analogy Gell-Mann drew was quite cogent, but irrelevant to the specific predictions he made from it! Gell-Mann had conflated two separate physical phenomena ("flavor" and "color") that happen to have the same symmetry! Thus, Gell-Mann introduced an extra, unintended analogy; the success of this latter analogy is unexplained even today.

[19] Ne'eman may have beat Gell-Mann to quarks; cf. Ne'eman, p. 581, and the references there cited.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ogy—there were powerful objections to particles, like quarks, with fractional charge: none had ever been observed. In other words, relevance, for me, is an *epistemic* concept—relevance is always relative to one's corpus of background assumptions.

My thesis, then, is: the use of formal mathematical analogies—not between events, but between their descriptions—has been indispensable in recent physical discovery. But such formal analogies appear to be irrelevant analogies—and irrelevant analogies should not work at all.[20] One would be quite surprised, for example, if one could make even one substantial physical discovery based upon the statistical distribution of the letters of the Roman alphabet in Newton's *Principia*. True, not all the mathematical analogies examined below reach this particular height of irrelevance—though some of them come pretty close. (I shall suggest, for example, that even the geometrical properties of the mathematical notation used in physics have played a role in physical discovery!) Relevance admits of degrees, and all my examples, I say, are sufficiently irrelevant that their success evokes, or should evoke, puzzlement. I arrive, then, at a Wigner-like conclusion: the success physicists have had in using mathematics to make discoveries is, indeed, "unreasonable." The role of mathematics in physical discovery thus raises a genuine, if overlooked, philosophical problem.

Some may be now inclined to protest that the very success of the mathematical methods of discovery described below demonstrates their relevance. My program, therefore, is self-defeating: the more examples of success in employing mathematical analogies I muster, the less surprising such success becomes.

The fallacy in this argument is worth attending to: a scientist, in employing a formal mathematical analogy, A, is relying not only upon A, but also upon the success of other formal mathematical analogies; that is, there is an analogy between A and other mathematical analogies. What property does A have in common with the others? The property, naturally, of being a formal mathematical analogy. We now must ask: Does the property, formal mathematical analogy, have a physical basis, at least on present lights? In other words, can we characterize physically the set of all mathematical properties? There is plenty of reason to believe the answer is: no, at least not now. If Wigner is right, and there are many mathematicians

[20] Relevance is relative, of course, to one's corpus of background assumptions; the judgments of relevance in this paragraph are relative to what I take to be widespread contemporary attitudes.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

who agree with him,[21] the class of mathematical concepts is defined primarily by aesthetic criteria—a positive reason to doubt that the notion of a mathematical concept can be characterized physically. Admittedly, for our purposes it would be sufficient to characterize physically the set of mathematical concepts applicable to physics. But, of course, no one today knows how to do this. Conclusion: buttressing a formal mathematical analogy by the success of other formal mathematical analogies is itself a formal mathematical analogy, just as irrelevant to physics as any other. What is more, there is no sense speaking about a general "phenomenon" of the applicability of mathematics to physics, so long as we have no way to characterize—physically—what we mean by 'mathematics', or by 'applicable mathematics'.

Similarly, our problem—the success of irrelevant mathematical analogies—cannot be dismissed by noting that many mathematical analogies were later grounded physically (quark theory was a case in point), and hence were, in fact, relevant. On the contrary, that is just our puzzle all over again: why formal analogies have turned out, so often, to have had physical content! Again, that formal analogies have, in the past, turned out to have had physical content is no reason to expect that present ones will, so long as the class of formal mathematical analogies, like the class of mathematical concepts, has no physical basis. Thus, each success of a formal mathematical analogy is an occasion for greater, not lesser, puzzlement.

But now we seem to be facing a paradox: this essay assembles the historical evidence for a hypothesis which, on standard scientific grounds, cannot be confirmed! There is nothing new about such a situation. Many philosophers of science, most prominently Nelson Goodman, have pointed to the inability of some generalities to be confirmed by their "instances." For example, it would have been superstitious not to run for the U.S. presidency in 1980, just because every twenty years, beginning in 1840, the president then elected died while in office. Of course, most of us do not ask for an explanation of this coincidence of U.S. history. The historical success of mathematics in physical discovery, on the other hand, does not seem to me to be a coincidence—and therefore demands some sort of account.[22]

[21] Von Neumann, in *The Mathematician:* "One expects a . . . mathematical theory not only to describe and classify in a simple and elegant way numerous and a priori disparate cases. One also expects 'elegance' in its 'architecture'. . . . These criteria are clearly those of any creative art. . . ."

[22] Compare the fitness of species to their environment. Darwin knew that this was no coincidence, yet could not simply accept it. Hence he sought to explain it, or

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

In this essay, I discuss only successful discoveries involving mathematical analogies; what of the many failures? By ignoring failures, one could make a case for astrology as a guide to political decision making. First, of course, mathematical analogies are only part—though a necessary part—of any discovery. Thus, when a scientist who has used a mathematical analogy fails to make a discovery, it is not necessarily the analogy that has failed. Similarly, mathematical analogies seem to succeed much more often in the hands of great physicists than in the hands of poor physicists.[23] Thus, the number of certified "failures" is much smaller than a superficial count might give us.

Second, even one successful use of irrelevant analogies to make major discoveries in physics is surprising. As we shall see, for example, Dirac discovered the relativistic wave equation for the electron with the aid of a formal mathematical analogy. Without the analogy, there was no serious possibility of discovering a relativistic equation correct to within experimental error (as of 1930). And, of course, no one could guarantee that there even was such an equation. When we combine this case with the many others discussed here, we have success that requires explanation. By contrast to the precision of a physical description, the "predictions" of astrology are sufficiently vague that we expect them to be "verified" on occasion.

Note, finally, that my thesis is much less vulnerable to the usual rebuttals—empiricist, Kantian, or Darwinian—than is Wigner's. Those rebuttals, if successful, would explain away the success of first-order mathematical analogies—why mathematical concepts have been so useful in describing and predicting the course of nature. They say nothing about using mathematical concepts in discovering those descriptions.

The evidence detailed here, therefore, persuades me that there is a genuine philosophical problem in the applicability of mathematics—one for the most part, and unjustly, ignored by contemporary philosophers.

II.

The procedures to be discussed are of five types:

(1) Equation E has been derived[24] under assumptions A. The

---

explain it away, as an artifact of natural selection. In this essay, I shall not attempt to formulate a criterion to distinguish between unreasonable hypotheses whose success is dismissed as coincidence and those whose success demands an account.

[23] This is not a tautology, since a physicist can show greatness in other ways than those described in this essay.

[24] By a 'derivation', physicists do not usually mean anything like a rigorous de-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

equation has solutions for which A are no longer valid; nevertheless, one looks for these solutions in nature, just because they are solutions of the same equation.[25] A solution of an equation is, of course, a mathematical description of a natural system; "being a solution of equation E" is a second-order property; hence we have here a second-order analogy.

This strategy guided Schroedinger's[26] discovery of wave mechanics. Schroedinger began with a pastiche of classical mechanics and optics—associating a particle or particles in an external field with a wave,[27] whose characteristics were derived from classical mechanics. This association was based upon two assumptions:

1. The dimensions of the "path" of the system are large in comparison to Planck's constant.
2. The external potential is nearly constant over time.

Differentiating, and relying on the two assumptions, Schroedinger obtained two equations: one in which the time did not appear, and the other in which the total energy did not appear. But now Schroedinger dropped the very assumptions that led to the equations![28]

Schroedinger knew very well what he was doing:

> The substitution of a *partial* differential equation for the equation of dynamics in atomic problems appears at first sight a very doubtful procedure, on account of the multitude of solutions that such an equation possesses (*ibid.*, p. 28).

Some of the "multitude of solutions," indeed, could be eliminated as "nonphysical"—for example, a wave function that explodes to infinity (in general, any function not "square integrable"). But what

---

duction from known principles, but plausible reasoning used to "write down" (discover) an equation. I use the term in the same sense.

[25] There are cases, though, where the scientist will rule out solutions as "nonphysical," as Einstein ruled out travel faster than the speed of light, though this is consistent with his equations of special relativity. Such cases are discussed in my "Events and Causality," this JOURNAL, LXXXIII, 5 (May 1986): 249–264.

[26] "Quantisation as a Problem of Proper Values," especially pts. II and IV, in *Collected Papers on Wave Mechanics*, J. F. Shearer and W. M. Deans, trans. (New York: Chelsea, 1978). It is a bit risky to draw conclusions about discovery from published papers. Nevertheless, a number of remarks in this second (and fourth) paper on wave mechanics indicate that Schroedinger put great stock in the reasoning—as distinguished from that of the first paper, the one usually cited in textbooks, which Schroedinger himself dismissed as "incomprehensible" (p. 13).

[27] Not a wave through physical space, but a wave through "configuration space."

[28] More accurately, he dropped the first assumption in pt. II and the second in pt. IV. The final form of the equation, the familiar equation with the complex coefficient, was only arrived at in pt. IV.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the remaining solutions had in common—in light of the known gulf between classical and quantum phenomena—was, virtually, only the Schroedinger equation. Thus, Schroedinger relied on a formal, second-order mathematical analogy. The procedure worked. The equation yielded known results as solutions, and previously unknown solutions were found in nature.

Maxwell's procedure in writing down his immortal equations provides another example of this strategy. Once the phenomenological laws of Faraday, Coulomb, and Ampere had been given differential form, Maxwell noted that they contradicted the conservation of electrical charge, though the phenomenological laws were strictly in accord with the evidence then available. Yet, by tinkering with Ampere's law, adding to it the "displacement current,"[29] Maxwell succeeded in getting the laws actually to imply charge conservation. With no other empirical warrant (Ampere's law stood up well experimentally; on the other hand, there was "very little experimental evidence"[30] for the physical existence of a "displacement current"), Maxwell changed Ampere's law to read that (the "curl" of) the magnetic field is given by the sum of the "real" current and the "displacement current." Ignoring the empirical basis for Ampere's law (magnetism is caused by an electric current), but by formal mathematical analogy, Maxwell now asserted the law even for a zero "real" current! Thus did Maxwell predict electromagnetic radiation, produced later by Herz.

(2) One looks for solutions even where there is reason to doubt the physical reality of the solutions. Naturally, this search must be circumscribed, because it is just as much part of physics and physical discovery to reject solutions deemed nonphysical as to look for them.[31] This strategy exemplifies, more than the previous one, formal mathematical analogies—we believe in the equation because it is a mathematical description!

Consider the development of relativistic quantum mechanics. The Dirac equation for relativistic quantum mechanics described the behavior of the electron and the hydrogen atom, beautifully, but it also described particles of negative energies, an apparent impossibility. It

[29] Ampere's law, as formulated by Maxwell, states that the "curl" of the magnetic field at any point is proportional to the current at that point. Maxwell added the "displacement current," an hypothetical current, equal to the time rate of change of the electric field.

[30] James Clerk Maxwell, *A Treatise on Electricity and Magnetism* (New York: Dover, 1954; reprinted from the 1891 Oxford ed.), 3rd ed., vol. II, sec. 608, p. 252.

[31] Recall Einstein's rejection of velocities greater than c, and Schroedinger's rejection of exploding solutions.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

is true that (relativistic) classical mechanics also has such solutions, but these are routinely discarded as nonphysical; in quantum mechanics, where strange things happen if you wait long enough, one needs an explanation why a particle cannot make a transition from positive to negative energy. Dirac began to realize that the Pauli "exclusion principle" would prevent electrons from sinking back into negative energy states, if they were already occupied by negative energy electrons, so ubiquitous as to be undetectable. An electron, raised to positive energy by some means, would leave behind a "hole," i.e., an unoccupied negative energy state. Such a "hole" would act like a positively charged electron (a "positron") of positive energy. In fact, positrons were discovered.

Consider next the history of the Schwarzschild solution for the equations of general relativity. This solution describes the gravitational field of a spherically symmetric body, and for bodies of "normal" size and mass, gives not only Newton's laws as a first approximation, but deviations from those laws (such as the precession of the orbit of Mercury) as a second approximation. Unfortunately, if the body is smaller than its "gravitational radius" (for a body the mass of the sun, about 3 km), one of the components of the metric tensor goes to infinity at a distance of the gravitational radius from the center of the body. This seemed to indicate that no such body could exist, as it would create a "singularity" in space/time. Nevertheless, scientists had enough faith in the equation to believe in the possibility of even this special case of the Schwarzschild solution; and, in 1954, Finkelstein argued that the singularity at the gravitational radius was not a real singularity of space/time, but a mere artifact of the coordinate system adopted. From a far distance, a spherical mass smaller than its own gravitational radius would appear as a "black hole" (an idea going back to Laplace[32] at least); a test particle falling toward the mass would seem to "us" to take infinitely long to get to the gravitational radius, but an observer on the particle would not notice anything special as he crossed the gravitational radius. At any rate, this set off a search for black holes.

A final point: our search for solutions even when they look physically unreal may involve reinterpreting the solution or the equation; a number of our examples here and elsewhere reflect this. Modern physics records equations that were accepted, to some degree at least, even before they were interpreted satisfactorily. Quantum mechanics is an extreme example of this. This faith in the formalism as

[32] Thanks to Barry Simon for pointing this out.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

being "wiser than we"[33] is what motivates higher-order mathematical analogies.

(3) Suppose we have effected a successful classification of a family of "objects" on the basis of a mathematical structure *S*. Then we project that this structure, or some related mathematical structure *T*, should be useful in classifying other families of objects, even if (a) structure *S* is not known to be equivalent to any physical property, and (b) the relationship between structures *S* and *T* is not known to be reducible to a physical relation. Note that, in the latter case, we reach third-order mathematical analogies.

This kind of reasoning has been rampant in elementary particle physics, where "symmetry considerations" have been responsible for some remarkable discoveries. Some of these considerations are really formal arguments from analogy, where one symmetry is derived from another.

According to the modern definition, an object has a symmetry if it is invariant under a group of "transformations." For example, a ball has rotational symmetry, since any rotation leaves it as it was. The sun's gravitational field has rotational symmetry, even though the orbits of the planets, being ellipses, do not.

The importance of symmetries in physics lies in their relationship to laws of conservation. Each symmetry of a physical system implies a law of conservation[34]—rotational symmetry, for example, implies the conservation of angular momentum. In fact, for the conservation of angular momentum, all that is necessary is that the system in question be invariant under "infinitesimal" rotations.

During the 1920s, the electron was discovered to have an "intrinsic" angular momentum, called "spin."[35] Electronic spin has the peculiarity that, no matter how we define 'up' and 'down' in space,[36] the spin of the electron upon measurement is always directed "up" or "down." Thus, the electron, insofar as its spin is concerned, is always found to be in one of two possible states. But quantum me-

[33] Hertz, p. 449.

[34] For continuous symmetries, this was proved by Emmy Noether.

[35] The electron cannot be regarded literally as rotating; rather, for some purposes, it behaves "as if" it were rotating. Furthermore, although the electron is invariant under "infinitesimal" rotations, it is not quite invariant under finite rotations; an electron that has been rotated 360 degrees "interferes" with an electron that has not been rotated.

[36] In principle, there are three independent quantities defining angular momentum. But (a generalization of) the uncertainty principle prevents us from knowing more than one of these quantities. A single axis, conventionally the "z axis," is therefore chosen, by reference to which we measure the one knowable component of spin.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

chanics also allows particles, or systems of particles, having n possible spin states, for every natural number n.

In 1932, Heisenberg conjectured that the proton and the neutron are two states of the same particle, "spinning" in opposite directions in a totally fictitious three-dimensional "space" with the same mathematical properties as ordinary space. The space had to be fictitious, since (unlike the situation with the "up-down" electronic states) one cannot turn a neutron into a proton by standing on one's head. Heisenberg reasoned that the nucleus of the atom is invariant under abstract "rotations" in this fictitious space, and that there had to be, therefore, a new conserved quantity, mathematically analogous to spin. This quantity is today called isospin, and its discovery launched nuclear physics. By extending the formal analogy to spin even further, Nicholas Kemmer reasoned in 1938 that there could be an isospin triplet (a particle or system of particles capable of being in one of three isospin states), and predicted the properties of the three *pions* nine years before the experimentalists were able to verify them.

Isospin symmetry was the first example in physics of a purely abstract, or nonperceptual, symmetry. It was isomorphic, mathematically, however, to the kind of rotational symmetry characteristic of classical systems.[37] The next step was the introduction, by analogy, of abstract symmetries whose relation to experience was increasingly tenuous.

Consider, for example, the discovery of "unitary spin,"[38] also called the "eightfold way" because unitary spin has eight "components."[39] This scheme, discovered independently by Gell-Mann and Ne'eman, aimed to incorporate a newly discovered conserved quantity—"hypercharge," not to be discussed here—and isospin into a "global symmetry."

Like isospin, the theory of unitary spin postulated that strongly interacting particles that are prima facie different may actually be different states of the same "thing." Thus, the strongly interacting particles may be divided into families whose members share a common unitary spin[40] but are nevertheless in different unitary spin

[37] This is strictly true, again, only if we are thinking of infinitesimal rotations.
[38] This term, introduced by Gell-Mann, is not used today.
[39] Isospin, like spin, has three components, since we can have spin in the direction of the x, the y, and the z axis—only one of which, by the uncertainty principle, can be measured at a time. Unitary spin adds five more components to these three, of which at most two can be simultaneously ascertained.
[40] The unitary spin of a system is defined by an ordered pair of numbers, giving the isospin and the hypercharge of the system.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

"states." Again, this classification flows mathematically from a symmetry property of strong interactions, namely, that they are invariant under a transformation, as in the case of isospin.

But there are also a number of interesting novelties in the case of unitary spin. First, the number of particles belonging to a unitary spin family is severely restricted. Unlike the case of isospin, where families can have n members, for any natural number n, unitary spin families can have 3, 8, 10, 27, . . . members—the series being determined by algebraic considerations we cannot discuss here. In fact, one of the most spectacular predictions Gell-Mann and Ne'eman made, that of the "omega-minus" particle, was done by noting that 9 known particles could be regarded as belonging to a "decuplet" of unitary spin, provided that the missing tenth particle existed.

Second, the eightfold-way scheme grouped together particles for which there was little empirical evidence that they were merely "different states of the same thing." For this to be true, strictly, the particles would have to have the same mass/energy, yet, for example, in the "octet" family allowed by the eightfold way, the heaviest particles were 50% heavier than the lightest (which, by the way, were the proton and neutron)![41] What actually turned out to be the case, that unitary symmetry is "broken" by a natural effect that disguises the underlying order, might have looked like an ad hoc attempt to save the theory. Thus, the discoverers were gripped by a very strong faith in the symmetry of the basic forces of the universe.

But most important for our essay is this basic difference between unitary spin and isospin (to say nothing of spin): though unitary spin has eight components, it is not a rotational symmetry at all—not even in an eight-dimensional space! Instead, unitary symmetry is invariance under a transformation of a three-dimensional complex space![42]

Where did that come from? We know that isospin (and spin) is invariance under ordinary three-dimensional rotations.[43] From classical mechanics, it was known that three-dimensional rotations can be represented also as transformations in a two-dimensional complex space.[44] In classical mechanics, like all uses of complex numbers, this

[41] Compare the isospin families: the mass difference between proton and neutron, for example, is one part in seven hundred!

[42] For details, cf. Gell-Mann and Ne'eman, eds., *The Eightfold Way* (New York: Benjamin, 1964), which contains the original papers.

[43] Better, infinitesimal rotations.

[44] Cf. Herbert Goldstein, *Classical Mechanics* (Reading: Addison-Wesley, 1950), ch. 4, for a lucid account. The correspondence is an isomorphism only in the case of infinitesimal rotations, otherwise it is a homomorphism.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

was merely a representational convenience. But what Gell-Mann did,[45] in effect, was to generalize the concept of a "rotation" to include transformations of complex space, and then generalize the complex two-dimensional "rotations" to three dimensions! The crucial point is that, unlike the case of two dimensions, three-dimensional complex "rotations" no longer correspond to rotations of any real space of any dimension. The link with perception has been totally snapped. We have, thus, a mathematical analogy—a third-order analogy. The eightfold way derives from the fact that these three-dimensional complex "rotations"[46] turn out to have eight different components, corresponding to eight independently conserved quantities.

Now, how did Gell-Mann[47] discover the eightfold way? In particular, was the analogy to isospin physical or formal? Had he known then about quarks, which are the "triplets" of unitary spin mentioned above, he could have developed the eightfold way by building up the strongly interacting particles from quarks, just as Heisenberg built up the nucleus from nucleons (which are the fundamental "doublets" of isospin). In other words, if we know in advance that strongly interacting particles are made of quarks, unitary symmetry is a plausible hypothesis; no argument by analogy is needed to discover it.

But in 1960 there were good reasons to deny quarks. Quarks have fractional charge; and particles of this type had never been observed. Indeed, even after Gell-Mann finally put forward quark theory, with the suggestion that they might be unobservable, he was attacked by Marxist physicists for "bourgeois idealism."[48] What the Marxists had in mind, presumably, were passages like these:

> . . . we construct a mathematical theory of the strongly interacting particles, which may or may not have anything to do with reality, find suitable algebraic relations that hold in the model, postulate their validity, and throw away the model. We may compare this process to a

[45] Redid, since it had already been done by mathematicians.
[46] As always, the reference is to "infinitesimal" rotations.
[47] What follows is based on Gell-Mann's recollections in "Particle Theory from S-matrix to Quarks," and correspondence with Ne'eman. I reiterate: I discuss here only Gell-Mann's reasoning, because Ne'eman's route to the eightfold way (described in the same volume, pp. 499–510) did not involve mathematical analogies.
[48] See Gell-Mann, op. cit., p. 494. These are the physicists under the leadership of Sakata who had championed a "triplet" model of their own, in which "materialistic" protons, neutrons, and lambda particles make up strongly interacting particles.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

method sometimes employed in French cuisine: a piece of pheasant meat is cooked between two slices of veal, which are then discarded.[49]

In fact, Gell-Mann was led to the eightfold way by a tortuous road, involving a third-order mathematical analogy that he drew without knowing it. Indeed, the discoveries that he correctly predicted were not the ones that properly flowed from his analogy!

Gell-Mann made his discovery in an attempt to generalize the "Yang-Mills" equations.[50] These equations describe a field, whose relation to isotopic spin is intended to be analogous to that between the electromagnetic field and electric charge. The field itself has "local isospin symmetry": its behavior is invariant under independent rotations of the isospin at every point in space/time! (Thus, for example, what is called a proton at one point may be called a neutron at another. This is a much stronger symmetry, therefore, than Heisenberg's discovery.) Gell-Mann saw[51] that Yang-Mills theory provides a recipe for writing a field equation, given merely the appropriate local symmetry; hence, to generalize the equation, one must generalize the concept of isospin. What Gell-Mann did, in effect, but without knowing it, was to characterize isospin rotations as a "Lie algebra"—a concept he reinvented for the occasion, but which had been known to mathematicians since the nineteenth century. He then (apparently by trial and error) began looking for Lie algebras extending isospin—again, oblivious to the solutions already obtained—but failed, not realizing that the first solution required eight components, as above.[52] Later, a mathematician at California Institute of Technology realized what Gell-Mann was looking for, and enlightened him on Lie algebras. It seems, then, that Gell-Mann proceeded by a formal mathematical analogy, though an analogy he understood but insufficiently. Ironically, the symmetry Gell-Mann discovered was the symmetry of a property (called "color" in quark theory) that is not responsible for the classification of the strongly interacting particles. ("Flavor" is the property responsible for that.) Luckily for Gell-Mann, and for science, color and flavor have the same symmetry group, a coincidence for which no explanation is known.

It might be argued that, even without an underlying physical base, the use of group theory in physical discovery is empirically justifi-

[49] Gell-Mann, "Vector and Axial Vector Currents," in *The Eightfold Way*, p. 198.

[50] C. N. Yang and R. L. Mills, "Conservation of Istopic Spin and Isotopic Gauge Invariance," *Physical Review*, XCVI (1954): 191–195.

[51] "From S-matrix to Quarks," p. 489.

[52] He stopped at seven, exhausted by bouts of wine drinking, (p. 489).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

able, because group theory is grounded upon symmetry, an empirical notion. But this is circular reasoning, since "to say that an object has a symmetry just means that it admits a transformation into itself, and the collection of all these symmetries is a group."[53] The groups invoked in the theory of elementary particles today express symmetries only in this question-begging sense; they do not express the symmetries of any physical body or macroscopic physical system.

I conjecture, however, that what reinforces the feeling that certain groups express natural "symmetries" is that the notation that the mathematicians use to denote the groups is symmetrical. Similarly, a good mathematical notation may often suggest, pictorially, its own extension. A beautiful example of this is provided by our very example of isospin and unitary spin symmetries. Consider the following three matrices:

$$\begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix} \qquad \begin{matrix} 0 & -i \\ i & 0 \end{matrix} \qquad \begin{matrix} 1 & 0 \\ 0 & -1 \end{matrix}$$

Called the "Pauli matrices," they are fundamental to understanding spin and isospin symmetry, because they represent instantaneous rotations about, respectively, the x, the y, and the z axis of space (in the case of spin) or isospin space (in the case of isospin). How they do this is irrelevant to our present purposes.[54]

Compare, now, the eight matrices, which, in the same manner, represent instantaneous "rotations"—the eight components of unitary spin. These are, in essence, what Gell-Mann was looking for, but at first could not find. (He knew what algebraic relations they must satisfy, but could not find them.) Note that the first three matrices are, essentially, the Pauli matrices, which is to be expected, because three of the eight "components" of "unitary spin" are the three isospin components.

$$\begin{matrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} \qquad \begin{matrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{matrix} \qquad \begin{matrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{matrix}$$

$$\begin{matrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{matrix} \qquad \begin{matrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{matrix} \qquad \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{matrix}$$

[53] Bruce Chandler and William Magnus, *The History of Combinatorial Group Theory: A Case Study in the History of Ideas* (New York: Springer, 1982), pp. 52/3.

[54] For details, cf. Goldstein.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

$$\begin{matrix} 0 & 0 & 0 & 1/\sqrt{3} & 0 & 0 \\ 0 & 0 & -i & 0 & 1/\sqrt{3} & 0 \\ 0 & i & 0 & 0 & 0 & -2/\sqrt{3} \end{matrix}$$

The two groups of matrices have striking perceptual symmetry,[55] though the transformations, and hence the symmetries, they express are not perceptual at all (in the case of the Pauli matrices, I refer only to their use in representing isospin symmetry). Furthermore, the perceptual symmetries of each group of matrices are obviously related. The "naturalness" of isospin and unitary spin, as well as the transition from one to the other, is thus apparent—even to the layman—simply because of the symmetry of the notation. True, Gell-Mann did not know these latter matrices, but I am persuaded that, once he discovered them, these considerations strengthened his belief in the validity of the eightfold way, at least subconsciously.[56]

If this explanation of (some) mathematical developments is sound, the inadequacy of "empiricist" accounts of the applicability of mathematics follows immediately. For example, Nagel's thesis that "[i]t is a reasonable hypothesis that pure mathematics in general is so often applicable, because the symbolic structures it studies are all suggested by the natural structures discovered in the flux of things," quoted above in fn. 9, overlooks the possibility that some of the "suggestive structures" that prompt the development of mathematics and thus physics come, not from direct experience, but from the symmetries of the notation we use to report experience.

So far we have been looking at solutions to given equations; now let us consider inductive methods for "guessing" the equations themselves, based on mathematical projection.

(4) One formulates equations by analogy to the mathematical form of other equations, even if little or no physical motivation exists for the analogy. By definition, such reasoning is by formal mathematical analogies.

The most striking historical instance of this pattern of induction is perhaps Einstein's derivation of the field equations of general relativity. His method was to set down three mathematical conditions that the equation should satisfy—but then he discovered a mathe-

[55] Note, however, that a change of basis would destroy this perceptual symmetry; so that even the symmetry of the notation is somewhat arbitrary!

[56] The eightfold way was far from being accepted immediately. There were even, at first, mistaken experimental results that contradicted it. At one point, therefore, Gell-Mann wavered.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

matical proof that essentially only one equation satisfied them! We are interested here in two of Einstein's conditions: that the equation should be a second-degree differential equation, and linear in the second derivatives. Where did he get these conditions?

Einstein himself says[57] that the conditions that no more than the second derivatives of the metric tensor should appear in the field equation, and that the equation be linear in the second derivatives, both were "naturally taken from Poisson's equation." (Poisson's equation is the nineteenth-century form of Newton's law of gravitation, which indeed involves only the second derivatives of the "gravitational potential," and those only linearly.) By 'taken from Poisson's equation', Einstein certainly does not mean: the requirement that Poisson's equation should be a limiting case of the *GR* equation yields the second-order, linear condition, since that requirement could be met by a higher-order equation, or one not linear in the second derivative. He means that the coordinates of the metric tensor "play the role" of the gravitational potential of the Poisson equation. But, as John C. Graves[58] points out, and as an inspection of Einstein's reasoning[59] clearly shows, this argument rests on a formal analogy and dimensional considerations.

Another example of derivation by formal analogy is the procedure of Heisenberg (with Born and Jordan) in deriving matrix mechanics.[60]

Heisenberg began with the classical Hamiltonian equations of mechanics, and substituted matrices for the variables appearing in that equation. In Hamiltonian mechanics, for each system we construct a function $H(q, p)$ of the coordinates q and the momenta p—typically a polynomial in the qs and ps. The function H is called the *Hamiltonian* of the system, and reflects its specifics: spherical symmetry, an inverse square law, etc. Now we can construct differential equations of the form

$$f(H) = d/dtF(q, p, t)$$

which give the time dependence of a function F of the coordinates, momenta, and the time, as the system sweeps deterministically

[57] *The Meaning of Relativity* (Princeton: University Press, 1974), p. 84.
[58] *The Conceptual Foundations of Contemporary Relativity Theory* (Cambridge: MIT, 1971), p. 178.
[59] *Ibid.*, pp. 79–81.
[60] Werner Heisenberg, *Zeitschrift für Physik*, XXXIII, 879 (1925); M. Born, W. Heisenberg, and P. Jordan, *Zeitschrift für Physik*, XXXV, 557 (1925).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

through time and space, as a function f of the Hamiltonian.[61] The functional f will depend on the function F we want to calculate, but in all cases will contain differential operators.[62] [To simplify matters, I shall further abbreviate any Hamiltonian equation of this type as $E(H)$.]

Now, Heisenberg's idea—as amplified by Born and Jordan—was as follows. Given any quantum system that would have, if a classical system, Hamiltonian H, and be described, therefore, by an equation $E(H)$, simply replace all the variables p, q in the equation by matrices, and all operations (addition, multiplication, differentiation) by corresponding matrix operations. The result is a matrix equation $E^*(H^*)$, and this is the quantum equation that governs the system!

Now, granted that quantum mechanics is a theory involving matrices, why must the laws of the theory have the same overall form as those of classical mechanics? One might argue: by a physical analogy.[63] For example, the atom was known to resemble the solar system: inverse square law of attraction, spherical symmetry, most of the mass concentrated in the center. Why should the equations not have the same form? Yet this argument would, at best, explain the transition from H, reflecting properties of the solar system, to H^*, reflecting the similar properties of the atom.[64] It would not explain the transition of the predicate (or operator) E to the exactly analogous E^*, giving us equation $E^*(H^*)$. Why should these have the same form? (After all, we know that quantum systems exhibit weird behavior—that is why we need new equations in the first place.)

Can the correspondence principle of Bohr help? The usual interpretation of the principle is that the classical equations are to be the limiting case of the quantum equations, as the systems described get larger and larger with respect to Planck's constant h—alternatively, as we formally allow h to approach zero. But, clearly, the correspondence principle—so interpreted—does not determine the form of the quantum equations that describe the nonclassical shenanigans.

[61] We think of the system as a single point in a multidimensional phase space, with one dimension for each coordinate, each momentum, and the time. For example, if we have a system of two mass points, we will need six space coordinates, six momentum coordinates, and time, to describe the system.

[62] Thus, if all we want is the time dependence of the second space coordinate, so that F is nothing but a trivial "projection" function, then all we need to do is take the partial derivative of H with respect to the second momentum coordinate. In general, the story is much more complicated, and involves the "Poisson bracket" of F with H. Cf. Goldstein, chs. 2, 7.

[63] As the readers of an earlier version of the present essay argued.

[64] And there is no real argument that H^*, under the circumstances, need have the exact overall form as H.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

There is, to be sure, an interpretation of the correspondence principle which is relevant to Heisenberg's procedure—Dirac's interpretation.[65] After rediscovering, independently, a number of the insights of Born and Jordan, and adding to them,[66] Dirac concludes:

> The correspondence between the quantum and classical theories lies not so much in the limiting agreement when $h \to 0$ as in the fact that the mathematical operations on the two theories obey in many cases the same laws (*op. cit.*, p. 315).

A more blatant formal mathematical analogy would be hard to find. And what of Heisenberg? It seems clear that Dirac's sentiment must have motivated Heisenberg's procedure, at least in part. What divided Dirac from Heisenberg, if anything, was the weight each gave to purely mathematical analogies in physics; both agreed, however, that mathematical analogies have some weight.

Another historical example of the derivation of an equation based on a mathematical analogy is that of the Klein-Gordon equation, which is a relativistic version of Schroedinger's equation—in fact, despite its name, Schroedinger himself derived it independently.[67] Schroedinger knew that his (nonrelativistic) equation could be obtained from the classical energy-momentum relation

(CE) $$E = p^2/2m$$

by substituting differential operators for E and p. Now the corresponding relativistic equation is

(RE) $$E^2 = c^2p^2 + m^2c^4$$

Schroedinger thus suggested making the identical substitution for E and p as in the classical case, and obtained the Klein-Gordon equation.

More precisely, Schroedinger did not write down the Klein-Gordon equation in this form, which is an equation for a free particle. For, according to that equation, the probability for the particle to be in a certain region can be negative! Instead, Schroedinger multiplied the probability density function by the electric charge e, interpreting it as a charge density function—charge density, unlike probability

[65] In "The Fundamental Equations of Quantum Mechanics," *Proceedings of The Royal Society*, A109 (1926): 642–653; reprinted in B. L. Van Der Waerden, ed., *Sources of Quantum Mechanics* (Amsterdam: North Holland, 1967), pp. 307–320.

[66] Dirac discovered the formal analogy between the classical Poisson brackets and the commutator of two operations or matrices in quantum mechanics.

[67] Schroedinger, pt. IV, pp. 118–120. Abraham Pais, in *Inward Bound* (New York: Oxford, 1986), lists six authors who derived this equation in the space of half a year, in 1926.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

density, can be negative.[68] And his equation was derived by substituting in the relativistic energy-momentum equation for an electron in an electromagnetic field, rather than in that of the free particle. What justified these substitutions? In Schroedinger's own words, nothing but a "purely formal analogy"![69]

True, Schroedinger expresses "the greatest possible reserve" about his relativistic equation (*ibid.*, p. 119). And one reason he gives is the lack of physical basis for his derivation. So one might suspect Schroedinger of rather weak faith in the applicability of mathematics. But his real objection, as he himself points out, is that the equation does not yield the correct energy levels of the hydrogen atom—these are influenced by the spin of the orbital electron, and the solutions of the Klein-Gordon equation do not have "spin." (The concept of spin can be added on an ad hoc basis to the nonrelativistic Schroedinger equation—but any attempt to do so to the Klein-Gordon equation destroys its Lorentz invariance.) Charged particles with no spin were not at that time known.

Another problem with the Schroedinger interpretation of the Klein-Gordon equation was the existence of "negative energy" solutions, just as in the Dirac equation. Dirac's treatment of negative energy solutions, described above, could not work here, because the Pauli exclusion principle is not true for spinless particles. It took until 1934 to "understand" the Klein-Gordon equation, when Pauli and Weisskopf interpreted it as a field equation rather than a particle equation. In another application of inductive pattern (4), they applied to the Klein-Gordon equation the so-called "field quantization technique," which had been used to extend Dirac's equation for the electron to a many particle theory, able to describe the creation and destruction of material particles. In the case of the Klein-Gordon equation, however, Pauli and Weisskopf began with an equation that had no physical solutions—and derived a field equation with, finally, a positive energy parameter. This equation[70] does, in fact, describe spinless, charged particles: for example, the positive and negative pions, discovered subsequently. Our story, then, doubly illustrates pattern (4): in the transition by analogy from Einstein's equation to

[68] This move itself illustrates pattern (2) above.

[69] Pp. 118/9. Schroedinger's notebooks reveal that he had derived, perhaps before the Schroedinger equation, a time-independent (thus, non-Lorenz-invariant) equation for a relativistic electron in a Coulomb potential. Contrary to common opinion, it is to this, not the Klein-Gordon equation, that Schroedinger refers in pt. I, in speaking of his failed attempts at a relativistic treatment of the hydrogen atom.

[70] I am referring to the complex Klein-Gordon equation.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the Klein-Gordon equation, and from the Klein-Gordon equation to its field-theoretic extension.

Indeed, the mathematical analogy of the Klein-Gordon equation to Einstein's equation was so compelling that Dirac, despite his rejection of the Klein-Gordon equation,[71] used it in deriving his own equation for the electron. Having derived the general form of his equation, but lacking the coefficients, Dirac pinned down the coefficients by requiring that any solution of his equation be a solution of the Klein-Gordon equation also. Whatever was wrong with the Klein-Gordon equation, he argued, it at least expressed the relativistic energy-momentum relation for the free particle—in quantum-mechanical form. Dirac soon discovered that there were no numerical coefficients for which his equation implied the Klein-Gordon equation. Instead of giving up the transequational analogy between classical magnitudes and quantum operators, Dirac introduced matrices—not, as Heisenberg, in place of the variables of mechanics, but as coefficients in his equation. The rest of the story is textbook stuff—matrices as coefficients required the wave function to quadruple itself. Nevertheless, each of the four components of a Dirac solution satisfied the Klein-Gordon equation, as required. And the introduction of matrices turned out to be a blessing, because now the phenomenon of electron spin simply fell out of the equation.[72]

Dirac's derivation illustrates another, final example of the use of a mathematical analogy in deriving an equation. Schroedinger had derived the Klein-Gordon equation from Einstein's relation; but Dirac began his derivation with the Schroedinger equation. This equation asserts that the time derivative of the wave function is proportional to the "Hamiltonian" applied to the wave function—the proportionality constant being imaginary! By the Hamiltonian here is meant the result of substituting in the classical Hamiltonian (see above) the usual differential operators. Thus, the Hamiltonian, though formally analogous to the classical one, became itself a differential operator. Now, what Dirac proposed was simply to equate the first time derivative of the wave function with (some multiple of) the relativistic Hamiltonian applied to the wave function. For relativ-

[71] Cf. Pais, pp. 288–292. Dirac objected most of all to the second time derivative in the Klein-Gordon equation; in sect. 17 of his *The Principles of Quantum Mechanics* (New York: Oxford, 1958; 4th ed.), Dirac presents a "proof" that quantum mechanics requires a first-order time derivative.

[72] Namely, a solution of the Dirac equation consists in a quadruple of functions, which give, respectively, the probability (density) of finding at a point, x, of space/time, an electron with spin up, a positron with spin down, an electron with spin down, a positron with spin up.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

istic systems have Hamiltonians, just as Newtonian systems do, and we can substitute differential operators in relativistic Hamiltonian, just as before. Here Dirac met an obstacle. The resulting equation was not relativistic (Lorentz invariant) as a whole. Dirac's procedure then was to modify his Hamiltonian, making the smallest change that would render the entire equation relativistic. This procedure, together with the already described introduction of matrices, finally hit the jackpot. I shall leave, as an exercise for the reader, the task of calculating how many mathematical analogies guided Dirac in guessing the equation.

Recall, next, the Yang-Mills equations, which we mentioned in the context of a discussion of the role of symmetries in physics. But we could also mention it here, as a classic example of the use of a mathematical analogy in deriving an equation. The analogy was to electromagnetism: their equations were to be related to isospin as the electrodynamical equations are to charge. It is true, of course, that isospin is related to charge—charge marks out the "direction" of isospin. Yet the Yang-Mills equations ignore the electromagnetic interaction with charge. When we consider the extremely salient fact that the primary analogy between electromagnetism and Yang-Mills theory was a speculative analogy between the abstract symmetries of the two kinds of field, we are forced to the conclusion that formal analogies guided Yang and Mills. (I shall have more to say about Yang-Mills theory below.)

Our last examples of theory construction by mathematical analogies without physical ground derive from complex-function theory, and involve the concept of an analytic continuation. A function is called *analytic* in a region of the complex plane if it can be represented in that region by one or more convergent power series. Equivalently, a function analytic in a region is one whose derivative exists in that region; this equivalence, which is highly nontrivial, holds only for complex functions, not for real functions, because the existence of the derivative of a complex function is a much stronger condition than the existence of the derivative in the real case.[73] This is the case because, in the complex case, the existence of the derivative at a point means that the differential quotient has not only to converge to a limit, but to converge to the same limit, regardless of the direction from which we approach the point.

Now, a complex power series, typically, does not converge everywhere in the complex plane (though it may), but within a certain

[73] It is even stronger than the existence of the nth derivative for every n.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

"circle of convergence." (Where the power series converges everywhere, the radius of its circle of convergence is said to be infinity.) Thus, if a function is represented in a region by one or more power series, it is not immediate that it can be extended to be an analytic function in wider regions. Remarkably, however, it can be shown that any such extension is unique—so that, if we begin representing an analytic function by a power series in its circle of convergence, we may be able to "continue" the function analytically by the method of overlapping circles. And, fortunately, the historically useful functions turn out to be analytically continuable from the real line to the entire complex plane, with the exception of isolated singularities [such as zero in log (x)].

The concept of analytic continuation is central in defining a concept that became important to physicists in the 1960s—"crossing symmetry."[74] It was discovered that the equations of quantum field theory imply that the interactions among elementary particles are described, probabilistically of course, by analytic[75] functions of such parameters as the energy of the interaction. More: the equations imply that the functions describing interactions of particles have "crossing symmetry." That is, if a function, e.g., $f(E)$,[76] describes a collision of particles at energy $+E$, then, if we analytically continue f to $-E$, $f(-E)$—which has no obvious physical meaning—nevertheless describes the expected results of a collision involving an antiparticle! For example, consider the function describing "Compton scattering," in which an electron hits a photon and both are affected. By applying crossing symmetry twice, we get the probability amplitude for the annihilation of an electron-positron pair to produce two gamma rays (photons)! Notice that the path that we pursue in extending the function from $+E$ to $-E$ need not stick to the real line; it can wander off into the imaginary realm in order to outflank singularities. Even on the real line, the function may be extended into regions that are "unphysical"—for example, where the energy of the system is less than its mass.[77]

So far, we have an absorbing mathematical property of quantum field theory. But now comes the induction: "all experimental evidence accumulated so far supports the view that crossing symmetry

[74] Cf. A. S. Wightman, "Analytic Functions and Elementary Particles," in *The Mathematical Sciences*, pp. 118–127.

[75] More precisely, piecewise analytic, cf. Wightman, p. 120.

[76] I am simplifying by leaving out other variables than energy. As Barry Simon has pointed out to me, it is quite important in crossing symmetry that one has more than one variable.

[77] Again, this point was impressed upon me by Simon.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

holds in Nature."[78] This prediction, made in 1969, was not merely a description of current equations; it was a constraint to be followed in constructing new theories, outside quantum field theory—just as "Lorentz invariance" is meant as a constraint on constructing new theories. Yet, unlike Lorentz invariance, a mathematical constraint based on a physical idea (relativity: equivalence of all inertial systems), the requirement of crossing symmetry as a feature of any physical theory of subatomic collisions has little physical basis.

If we limit ourselves to quantum electrodynamics, the "electroweak" theory, and any other relativistic theory that obeys the principles of "locality" and "causality"[79] there is, in fact, a simple argument for crossing symmetry. Suppose that a collision occurs at a given point ("locality"). Then, in classical special relativity, none of the scattered particles can travel faster than the speed of light; otherwise, some observers would see the scattering before the collision, violating causality. In quantum theory, however, if a collision is localized, the uncertainty principle tells us that there is a finite probability to find one of the particles outside of the "light cone," i.e., having traveled faster than the speed of light! Crossing symmetry resolves this contradiction. The behavior of a positron is exactly that of an electron moving backward in time! (And so, of course, for all antiparticles.) Thus, the observer sees, not particles moving backward in time, but antiparticles moving normally.

In cases such as these, the concept of analytic continuation plays little role. We have a general description into which we can substitute various situations—hence, it is not in any way mysterious that two state descriptions are analytic continuations of each other.

But there are theories (such as so-called "*S*-matrix theories") to which this argument does not apply. These theories do not describe the progress through space/time of a "wave function" (or its generalization in field theory) from which predictions are extracted. Rather, the theories predict the results of scattering directly; no principle of locality is assumed. Furthermore, there is no general scheme into which we can "substitute" to yield scattering amplitudes for various kinds of event. To the contrary, crossing symmetry is a constraint built into the theory on the basis of a formal mathematical analogy. For example, the present rage, "string theory," violates locality, at least at present—yet the theory is constructed to obey crossing symmetry.[80] Admittedly, string theory has its detractors, so the success of this particular analogy cannot yet be finally assessed.

[78] Wightman, p. 124.

[79] Locality is the ability to localize an event at a point in space/time; causality is the principle that the cause precedes its effects.

[80] I am not claiming that crossing symmetry is the major idea in string theory,

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The conception that the analytic continuation of an important function is also important is not limited to the phenomenon of crossing symmetry. For example, it has recently been discovered that, if we "rotate" the complex plane 90 degrees, the basic formulas of quantum field theory are transformed into the basic formulas of statistical mechanics. This fact makes calculations possible which previously were impossible (instead of direct calculations in one theory, one "rotates" the problem, calculates in the other theory, then "rotates" back). It also has led to predictions that analytic continuations will lead to new discoveries and insight into old ones.[81] Since the concept of "analytic continuation" is, at present, a purely mathematical concept, any such prediction, even if based on past successes, is still a formal mathematical analogy.

Our final example of a mathematical analogy based on an analytic continuation leads, not to a new equation, but a new kind of calculation. Here is a simplified example.[82]

Problem: to calculate the "critical exponent" of a ferromagnet—how fast it loses its magnetism above a "critical" temperature. To calculate this directly from first principles[83] is hopeless. Remarkably, however, the corresponding problem for a "four- (or higher) dimensional magnet" is solvable by various approximation techniques. Wilson let the solution be an analytic function of the number d of dimensions (so that we can speak of "dimension 3.7239" or even "dimension $3.5 + .1i$") and continued the solution back down to dimension 3! (In some cases of this type of procedure, singularities blocking the way from 4 to 3 have to be circumvented, so that the function veers off the real line entirely!) The answer was amazingly accurate.

This procedure would be deductively legitimate, however nonsensical the physical motivation, if Wilson could have made the following assumptions:

(a) We have an "algorithm," based on first principles, defining an analytic function $f(d)$ which for $d = 3, 4, \ldots$ gives the critical exponent of a magnet of that dimension.
(b) Around $d = 4$, we can expand $f(d)$ by a power series $s(d)$.

---

merely that it is one constraint on string theory, one which at present has no physical basis.

[81] See Manin, pp. 80–84.

[82] K. G. Wilson, *Physical Letters,* XXVIII, 548 (1972).

[83] Wilson won the Nobel Prize for applying mathematical methods of quantum field theory to thermodynamics—a possibility already alluded to above. What is described here was part of that effort.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

What were the facts?[84]

(a) Based on first principles, Wilson could not prove mathematically that $f(d)$ makes sense even for natural numbers. (There is computer and other evidence, however, that it does make sense for natural numbers.)

(b) A fortiori, there was no mathematical proof that $f(d)$ was defined as an analytic function of dimension.[85]

(c) There was no mathematical proof that the actual series $s(d)$ Wilson used is summable; it seems not to be, in fact.

(d) Even if both f and s were analytic, it would not follow that s could be used in place of f, unless s is truly the expansion of f. Even if s coincided with the expansion of f, at $d = 4$, for the first 1,000 (or any finite number of) terms, its analytic continuation could differ drastically from f at $d = 3$![86]

What, then, motivated Wilson's procedure, aside from blind faith? Well, the notion that the dimension can be a variable had been used elsewhere, in quantum electrodynamics.[87]

We seem driven to the conclusion, therefore, that Wilson, at a critical phase of his calculation, engaged in doubly analogical behavior: Wilson manipulated symbols in ways analogous to deductive mathematics, without actually engaging in mathematical deduction —justifying himself by analogy to other, similar, successful violations of mathematical rigor!

Now, physics has employed, and still employs, nonrigorous mathematics:[88] Newton's "fluxions," Dirac's "delta function," and Feynman's "path integrals" come to mind. But at least these procedures were intuitively reasonable prior to their rigorous formalization.[89] Wilson's "calculations" resemble, rather, Euler's formal (but highly successful) manipulation of divergent series; for no physical intuition then backed the assumption that $f(d)$ is analytic, and that $s(d)$ can be

[84] I rely here on conversations with several physicists and mathematicians, especially Shmuel Elizur and Barry Simon. They are in no way responsible for any distortions.

[85] Elizur points out that f can be represented as a power series $t(g)$ in the "coupling constant," g, where each of the coefficients is a provably analytic function of d. Alas, $t(g)$ does not sum.

[86] Zalcman noted that I am speaking here of cases where the radius of convergence of s is less than unity, so that the analytic continuation is nontrivial.

[87] Joel Gersten points out that, for a two-dimensional magnet, the answer can actually be calculated directly, and thus checked against Wilson's procedure, giving added "inductive" support.

[88] I mean, of course, nonrigorous when first used.

[89] Berkeley's point was merely that the usefulness of the calculus could be explained without assuming its consistency.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

regarded as its expansion. After all, for example, $f(d)$ has (today) no physical meaning—if only because real (to say nothing of complex) dimensions have no physical meaning. Wilson had, then, neither rigorous mathematics nor physical intuition to support him; only a formal analogy.[90]

(5) One formulates basic laws using mathematical ideas which in earlier theories were mere calculational devices. This is true even if, by doing so, we grant these mathematical ideas "physical reality,"[91] which they did not have before. Nevertheless, our theory construction proceeds a pace, governed by mathematical analogy to the earlier theories—and, also, by analogy to other cases in which calculational devices became physically real.

For an example, take the so-called "vector potential" of classical electromagnetism. The vector potential[92] gives us the magnetic field by differentiation, just as the ordinary, "scalar" potential gives us the electric field by differentiation[93]—both potentials, but especially the analogous vector potential, are true mathematical conveniences. It is simply easier to solve equations for the potentials, and then find the fields by differentiation. That the potentials are not "real" can be seen by the so-called "gauge invariance" Maxwell's equations exhibit. What this means, roughly, is that, since the potentials are given by integrating the fields, we introduce extraneous "integration constants" into the definition of the potentials—called gauges. This makes the exact value of the potentials unimportant to observation. In fact, adding the partial derivatives of an arbitrary function to the potentials leads to the same equations of motion. Such a transformation of the potentials is called a "gauge transformation"; the equations of classical electromagnetism are thus "gauge-invariant."

In quantum mechanics, the potentials themselves appear in the fundamental laws.[94] Thus, the potentials could influence the course of an experiment in a region, even if the electromagnetic field there

[90] The analogy, though certainly formal, hardly strikes me as mathematical; "pseudo-mathematical" would be a better term. Needless to say, I mean here no disparagement; you go with what you have got.

[91] There are several criteria for physical reality in the literature, cf. Ernest Nagel, *The Structure of Science* (New York: Harcourt, Brace & World, 1961), ch. 6; no single one is meant here. As Nagel points out, moreover, criteria for physical reality are internal to physical theorizing itself—so that employing them does not commit the philosopher to metaphysical realism.

[92] So called because it has three components and behaves like a vector.

[93] The electric field is the "gradient" of the scalar potential, the magnetic field is the "curl" of the vector potential.

[94] In classical mechanics, one can write equivalent equations using either potentials or fields; not so in quantum mechanics, so far as is known.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

is zero! Aharonov and Bohm[95] suggested a two-slit experiment, in which, as usual, electrons slide through two holes and hit a screen, creating a pattern of interference. In this experiment, however, a solenoid is switched on between holes and screen. Outside the solenoid, the electromagnetic field is zero, but the potentials are not. Sure enough, a phase shift of the interference is observed.

Admittedly, quantum mechanics is also gauge invariant—just as before, the exact value of the potentials makes no physical difference. The Aharonov-Bohm effect depends, however, not on the value of the potentials, but on the result of integrating the potential on a closed loop around the solenoid. This magnitude is gauge invariant, and can indeed affect the result of the experiment—but only in quantum mechanics. Classically, a zero field leads to zero effects; for example, integrating the field around the same loop will still give zero. Potentials, therefore, since they can be used to construct entirely new (gauge invariant) quantities, are no longer just a mathematically convenient way to talk about fields. Despite this clear change of role, scientists formulated quantum mechanics in terms of potentials, in formal analogy to their use in classical theories.[96]

There is a geometrical way to discuss potentials which is inspired by Einstein's theory of gravitation, and which has led to further discoveries. In Einstein's theory, "turning on" a gravitational field leads to a distortion of the geometry of space/time. In the Aharonov-Bohm effect, the solenoid also creates a kind of curvature, though in this case a curvature of an abstract space called a "fiber bundle." The amount of curvature is measured by our friends, the potentials—just as the so-called Christoffel symbols in Einstein's theory measure the distortion of space/time away from flatness. And the quantum electron "feels" the curvature in its abstract space. Furthermore, just as Einstein's theory allows us to choose a different coordinate system at each point in space/time, without changing the laws of nature, so does a gauge transformation, in effect, "rotate" the coordinates of the abstract quantum space—differently for each point in space/time. Thus, both Einstein's theory and quantum electrodynamics have a local symmetry. The Aharonov-Bohm effect, being gauge-invariant, confirms that no change of coordinates can transform away the distortion of the fiber bundle caused by the

[95] *Physics Review*, LIX, 115, p. 485.

[96] More precisely, there are two equivalent ways to formulate classical theories—one uses fields, the other potentials. The latter formulation receives its legitimacy by its equivalence to the former. In the transition to quantum physics, however, the link between the field formulation and the potential formulation is snapped, because only the latter formulation is generalizable to Schroedinger's equation.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

solenoid, just as no change of coordinates can make a macroscopic gravitational field disappear. That the potentials themselves are not gauge invariant simply means, in quantum mechanics, that their numerical value is affected by an arbitrary choice of coordinates—much as the numerical value of length is affected by changing from feet to meters—not that the potentials are "unreal." To the contrary, potentials contain real information about the distortion of the fiber bundle, which is why we can use the potentials to define gauge invariant quantities that affect experiments.

This geometrical interpretation of the potentials of quantum electrodynamics, inspired by general relativity, was another source of the mathematical analogy that led to the Yang-Mills theory, already discussed twice.[97] Yang-Mills allows, for each point in space/time, a different choice of coordinates (equivalently, a different "rotation") in isospin space. Ironically, the Aharonov-Bohm effect can be predicted using quite elementary quantum physics; yet it was not discovered until about 1960, six years after the Yang-Mills theory had been published.

Another, even more basic, concept of quantum mechanics, previously a computational convenience, is that of the imaginary numbers themselves. Because of the formula

$$e^{ik} = \cos(k) + i \sin(k)$$

the expression e^{ik} was used in place of either sin or cos in representing periodic phenomena like waves. The convenience here is that the derivative of e^{ik} is (a multiple of) itself. Now, if we represent, say, $\cos(k)$ by e^{ik}, we "carry" along the "imaginary part" $i \sin(k)$, but—in second-order, linear differential equations (like Maxwell's)—this does not matter. For if e^{ik} solves such an equation, so does $\cos(k)$. Hence, we do the calculations with complex numbers, and then "separate out" the "real" part of the solution as "physically real."

In Schroedinger's equation, however, this convenience becomes a necessity. Both the equation and the solutions are irreducibly complex. This is connected with the derivation of the equation, discussed above: by part IV of his immortal essay on wave mechanics, Schroedinger has concluded that one can derive the correct wave equation only by taking a solution of the form

$$e^{ibt},\ b \text{ a constant}$$

[97] Hermann Weyl, earlier, had hoped that he could apply the concept of a gauge symmetry directly to general relativity, but nothing came of this.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

and differentiating it once, so that the resulting equation is first-order in the time. Such an equation no longer permits the separation of the solutions into real part/imaginary part. The solution is a complex function, period; complex numbers can no longer be regarded as calculational conveniences. (How to interpret these solutions was a problem solved, if at all, only with great difficulty.) Schroedinger, therefore, was relying on formal mathematical analogies between classical and quantum wave functions.

Let us here recall our earlier discussion of the use of two-dimensional complex matrices[98] to represent real rotations in three dimensions, a mere convenience in classical mechanics. Gell-Mann and Ne'eman generalized the complex matrices into higher dimensions, where the correspondence to real rotations no longer exists. Thus, the complex matrices can no longer be regarded as conveniences.

The discovery of the quarks is a final illustration of a calculational device turned "real." We have already noted the objections to the reality of quarks. On the other hand, quarks served as a handy device to classify the strongly interacting particles. The success of this formal, notational "construction" led physicists to expect that it was analogous to a real, physical construction of particles out of quarks, for which there is today considerable physical evidence.

There is ample evidence, then, that formal mathematical analogies lie at the heart of modern physical discovery. I hope that the reader agrees with me that an anomaly worthy of the philosopher's scrutiny confronts us.

MARK STEINER

Hebrew University/Jerusalem

[98] I refer, of course, to unitary matrices with determinant +1, the group known as $SU(2)$.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ANOTHER ARGUMENT AGAINST VAGUE OBJECTS*

THE view that the world might actually *be* vague, as opposed to being vaguely described and as opposed to not having sufficient information to determine whether some proposition is true or false, has not received a favorable reception. Bertrand Russell[1] claims that it "is a case of the fallacy of verbalism—the fallacy that consists in mistaking the properties of words for the properties of things"; Michael Dummett[2] says that "the notion that things might actually *be* vague, as well as being vaguely described, is not properly intelligible"; and Gareth Evans[3] presents a logic-based argument against it. Despite the fact that most agree with the anti-vagueness-in-reality conclusion, the argument has generated considerable criticism; but the criticism has been diffuse in the sense that the commentators do not always agree on the very structure of the argument nor on the possible objections to Evans's assumptions. Evans said (I use λx rather than $\hat{x}$ for abstraction):

> Let a and b be singular terms such that the sentence '$a = b$' is of indeterminate truth value, and let us allow for the expression of the idea of indeterminacy by the sentential operator, '∇'. Then we have:
> (1) $\nabla\ (a = b)$
> (1) reports a fact about b which we express by ascribing to b the property '$\lambda x[\nabla\ (x = a)]$'
> (2) $\lambda x[\nabla\ (x = a)]b$
> But we have
> (3) $\sim\nabla\ (a = a)$
> and hence:
> (4) $\sim\lambda x[\nabla\ (x = a)]a$
> But by Leibniz's law, we may derive from (2) and (4):
> (5) $\sim\ (a = b)$

* I am grateful to a number of people for discussions about this paper or topic: Bill Blackburn, Ron Clark, Matthew Dryer, Michael Dunn, Andre Fuhrmann, Randy Goebel, David Lewis, Bernard Linsky, Terry Parsons, Graham Priest, Len Schubert, Richard Sylvan, Alasdair Urquhart, and Johan van Bentham. I wish also to thank Zorba McRobbie and Frank Jackson of the Australian National University for providing a pleasant environment in which to think about vagueness during 1987–88.

[1] "Vagueness," *Australasian Journal of Psychology and Philosophy*, I, 2 (1923): 84–92.

[2] "Wang's Paradox," *Synthèse*, XXX, 3/4 (1975): 301–324.

[3] "Can There Be Value Objects?" *Analysis*, XXXVIII, 4 (1978): 208. N. Salmon, *Reference and Essence* (Princeton: University Press, 1979), pp. 243ff, esp. fn. 24, produces a proof, somewhat different from Evans's, to the same conclusion (that vagueness of objects is impossible). He also thinks that Evans's proof is correct. These arguments have been widely discussed in the literature.

0022-362X/89/8609/481–492 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

contradicting the assumption, with which we began, that the identity statement '$a = b$' is of indeterminate truth value.

Now *this* argument, phrased in terms of "truth values," would seem to read ∇ as meaning "has some truth value other than 0 or 1." A little later, however, Evans remarks, "if Δ [the dual of ∇, meaning 'is definite'—FJP] determines a logic at least as strong as S_5," then we can further derive $\Delta\ (a \neq b)$ from the supposition that $\nabla\ (a = b)$. So the question arises: Does Evans intend his Δ and ∇ to be modal operators or "truth-value indicating" operators? In either case—and our commentators are divided on which he intended—there seem to be difficulties with the argument.

It is my opinion that these two understandings of ∇ and Δ correspond to two radically different conceptions of that to which vagueness is due. A modal interpretation of vagueness takes each statement ϕ as being either true or false at any given index (world), but claims that prefixing, say ∇, to such a statement orders one, in effect, to look at some related (accessible) index to determine whether ϕ is true or false there before $\nabla\phi$ can be judged true or false at the initial index. Such an outlook does not seem to make $\nabla\phi$ say that ϕ is vague "in a realistic sense," or that ϕ is vague in the sense of "the world being vague"—after all, by hypothesis, ϕ *is* either true or false. Rather, such an account would seem most happily to find a home as an account of epistemic vagueness: ϕ is either true or false, but we do not know which, either because we lack the information necessary to determine it or because we are not clear as to which proposition ϕ expresses. But it nonetheless does have a truth value. Contrasted to this, a *truth-value interpretation* of $\nabla\phi$ actually says that ϕ is neither true nor false, and therefore that "the world" takes on this indeterminacy. It should also be recalled that Evans is explicitly interested in the claim that "the world might actually *be* vague, as well as being vaguely described." So, the truth-value interpretation of Δ and ∇ would appear to be more suited to his overall aim than a modal-operator interpretation (despite what were perhaps his own intuitions[4]).

There are a number of principles concerning ∇ and Δ which one might consider reasonable in interpreting them as "definite" and

[4] David Lewis informs me that Evans had written him a letter confirming that ∇ and Δ were intended as modal operators, and that it was a slip when, in the paper, he mentioned that p followed from Δp; for p is allowed to be definitely false as a way of making Δp true.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

"indefinite" operators.[5] Evans himself mentions only two basic ones, (*duals*) and (*refl*), but also seems to suppose that definiteness and indefiniteness are contraries, (*contra*):

(duals) $\nabla\phi = df \sim \Delta \sim \phi$ (that ∇ and Δ are *duals*)

(refl) $\Delta\ (\alpha = \alpha)$, for all α. (the reflexivity of identity is definite)

(contra) $\Delta\phi \leftrightarrow \sim\nabla\phi$ (that no sentence is both definite and indefinite)

Whichever of the two interpretations one takes of Δ and ∇, the structure of the argument is:

suppose (i) $\nabla\ (a = b)$

then (ii) $\lambda x[\nabla\ (a = x)]b$

equivalent to (i) by lambda abstraction

but (iii) $\sim\lambda x[\nabla\ (a = x)]a$

from (refl) by lambda abstraction and (contra)

ergo (iv) $a \neq b$ from (ii)–(iii) Leibniz's law

One objection to Evans's argument is that there is no contradiction here: $\nabla\ (a = b)$ can happily coexist with $a \neq b$, contrary to Evans's assertion that this is a "contradicting [of] the assumption, with which we began, that the identity statement '$a = b$' is . . . indeterminate. . . ." At best, or so it is claimed (see John Heintz and Peter van Inwagen; the point is implicit also in John Broome and Terrance Parsons[6]), the argument proves

(v) $\nabla\ (a = b) \rightarrow a \neq b$ from (i)–(iv) by conditional proof

Perhaps, but it seems not to have been also noted that from this we get the following:

(vi) $a = b \rightarrow \sim\nabla\ (a = b)$ contrapositive of (v)

(vii) $a = b \rightarrow \Delta\ (a \neq b)$ (vi) and (duals)

So, if in fact a and b are identical, then they are definitely distinct. If one holds the principle

[5] For a survey, see my "The Not-So-Strange Modal Logic of Indeterminacy," *Logique et Analyse*, XXVII, 108 (1984): 415–422.

[6] Heintz, "Might There be Vague Objects?", paper read to Canadian Phil. Assn, Halifax (1981); van Inwagen, "How to Reason About Vague Objects," *Philosophical Topics*, XVI, 1 (1988): 255–284; Broome, "Indefiniteness in Identity," *Analysis*, XLIV, 1 (1984): 6–12; and Parsons, "Entities Without Identity," *Philosophical Perspectives*, I, 1 (1988): 1–19.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(T) $$\Delta\phi \rightarrow \phi$$

(as P. F. Gibbins[7] and Broome do), then we can infer from (vii) and (T)—letting ϕ be $a \neq b$

(viii) $$a = b \rightarrow a \neq b$$

from which the troublesome

(ix) $$a \neq b$$

follows. But this result holds for any a and b (recall that it is no longer conditional upon $\nabla\ (a = b)$), in particular it holds for a and a:

(x) $$a \neq a$$

which everyone thinks is contradictory. (And, if not, then from (refl) and (T), we could derive an even more explicit contradiction.) It seems to me that this shows that allowing (i)–(iv) to stand already shows that the believer in vagueness-in-reality is in trouble. What is needed is to attack directly the (i)–(iv) argument.

Surely the most common attack is to focus on the use of lambda abstraction, perhaps in conjunction with Leibniz's law. This line of attack on the argument takes one of two forms. The first form points out that, generally speaking, one cannot "substitute into" so-called opaque contexts, and that treating ∇ and Δ as modal operators is as likely to generate such contexts as any other modal operators. Surely, so it is argued, we should not expect to say that a has all the modal properties that b does, even when '$a = b$' is true, unless the names a and b are "rigid designators" (which are defined so as to have all modal properties in common, if they are identical). This subline of attack locates the difficulty with Evans's argument in its passage from steps (ii) and (iii) to step (iv); or, more precisely, in the statement of Leibniz's law—the proper statement of the law would say (in its contrapositive form): "if property ϕ is predicated of the singular term 'a' and the property $\sim\phi$ of the singular term 'b', then if 'a' and 'b' are both rigid designators, it follows that $a \neq b$." And the correlative claim about vague identity would be that the only identity statements that can be vague are those wherein at least one of the singular terms is not rigid. So, on this line of thought, what Evans's argument proves is that you cannot have statements of vague identity when both terms are rigid. An approach similar to this for the modal

[7] "The Strange Modal Logic of Indeterminacy," *Logique et Analyse*, XXV, 100 (1982): 443–446.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

argument is given by David Wiggins,[8] where the point is put in terms of the scope of definite-description operators with regards to ∇ and Δ.

The second form this objection might take is to deny that every lambda expression designates a ("real") property; in particular, it is claimed that not all modalized formulas designate a "real" property. Leibniz's law is then stated in its second-order form: $a = b \leftrightarrow (\forall F)(Fa \leftrightarrow Fb)$, where the quantifier ranges over "real" properties. According to this view, two objects are distinct just in case they disagree on a "real" property. And, to complete the picture, lambda abstracts formed with Δ and ∇ do not designate "real" properties; therefore, Evans's argument goes awry (see Richmond Thomason[9] and van Inwagen). Where does it go awry? There are two ways to answer the question: one could say that such lambda abstracts are ill-formed, since they do not indicate "real" properties [this would be to deny the legitimacy of moving from (i) to (ii) and of moving from (refl) to (iii)]; or one could allow their well-formedness, but deny that Leibniz's law holds in such cases [here one denies that (iv) can follow from (ii) and (iii); for actually to apply Leibniz's law, one must first prove that the abstracts or predicates in question "really" do designate properties].

However plausible this rebuttal to Evans's argument might seem on first reading, it owes all its force to the analogy with modal logic and related concepts ("opaque contexts," "modalized properties," "real properties," "rigid designation," etc.). But this is not the only way to understand Evans's argument; indeed, as I remarked before, it is not even a correct way to look at it, if the argument is to be taken as being directed against "vagueness in reality." For the believer in vagueness-in-reality is committed to a many-valued logic, and therefore the Δ and ∇ operators must be understood as truth-value indicating operators. This is the only way of viewing the logic as "talking

[8] "On Singling Out An Object," in Philip Pettit and John McDowell, *Subject, Thought and Context* (New York: Oxford, 1986), pp. 171–182, esp. pp. 176/7; see also H. Noonan, "Vague Objects," *Analysis*, XLII, 2 (1982): 3–6, where he talks of a singular term "vaguely denoting" an object. Wiggins takes Evans's argument to prove something different from (v), viz., $a = b \rightarrow \Delta\ (a = b)$. This becomes our (v) if it is assumed that a statement is definite just in case its negation is. About this conclusion, Wiggins says it shows that "there is no future in the supposition that one could say that a was b but refuse to affirm that it was definitely b" (p. 175). It is not clear to me how Wiggins thinks he has reconstructed Evans's argument here, except perhaps by ignoring the requirement that we start with the supposition that '$a = b$' is vague.

[9] "Identity and Vagueness," *Philosophical Studies*, XLII, 3 (1982): 329–332.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

directly about reality" rather than about our statements or our knowledge. In reconstructing this argument—Evans's real argument against vagueness-in-reality, we will avoid using Δ and ∇ (because of their connections to modal logic—see the series of papers in the late 1960s in *Logique et Analyse* by Hugh Montgomery and Richard Routley where these symbols were used as "contingency" and "non-contingency" operators). Instead we will use the more common J operators.[10] For each truth value i, there is a sentence operator J_i which, when applied to a sentence, has the effect of claiming that the sentence has exactly the truth value i. If the value of ϕ is i, then $J_i\phi$ takes the value 1 ("completely true"); otherwise, $J_i\phi$ takes the value n ("completely false"). From these basic operators, more complex ones (e.g., "has at most the value i," or, "is either 'most true' or 'most false'," etc.) can be defined, should one wish.[11] Since we will concentrate at the outset on a three-valued logic, we have $J_1\phi$, $J_2\phi$, and $J_3\phi$. If ϕ takes the value 2, for example, then $J_2\phi$ takes the value 1 and each of $J_1\phi$ and $J_3\phi$ takes the value 3. The intuition here is that $J_1\phi$ means that ϕ is definitely true, $J_2\phi$ means that ϕ is indeterminate, and $J_3\phi$ means that ϕ is definitely false. In such a setting, Evans's argument is:

suppose	(i′)	$J_2(a = b)$
then	(ii′)	$\lambda x[J_2(a = x)]b$
but	(iii′)	$\sim\lambda x[J_2(a = x)]a$
ergo	(iv′)	$a \neq b$

Under the assumption that conditional proof works as usual, we would then derive

(v′) $$J_2(a = b) \rightarrow a \neq b$$

analogously with before. And, as before, the choices seem to be either to deny that λ abstraction is legitimate or to deny that Leibniz's law applies to such predicates. Another option comes with nonclassical truth values, however: deny the validity in general of Leibniz's law. To the person who wishes to deny the legitimacy of λ

[10] See B. Rosser and A. Turquette, *Many-Valued Logics* (Amsterdam: North Holland, 1952).

[11] Note that such definitions can be used in the other direction also. If the language has, for example, constants for each of the truth values and a connective that operates like our $\leftrightarrow$ (below), then the J-operators can be defined. Indeed, there are many different ways to be able to have J-operators in the language by such definitions. I merely insist that the language be rich enough to "talk about" vagueness—so it must have the resources to define the J-operators.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

abstraction, it should be pointed out that the present logic is completely extensional. Each formula, and its parts, is evaluated "in the actual world." (That is, to find out the semantic value of any given formula in a given row of a truth table, one need only look at that row, not at any other rows.) This means that the grounds for denying λ abstraction which might be used in the modal case are taken away. Among other things, it follows that one can "substitute into" any part of any formula and preserve truth, regardless of whether it has λ abstracts over J operators. In general, if a and b are names in this language and '$a = b$' is true, then any predicate that can be formulated in the language will apply truly to a just in case it applies truly to b. This is part of what is meant by saying that the language is truth-functional. There are no "rigid designators," no "modal properties," no distinction between "predicates that do and predicates that do not designate real properties," or the like. So this evasion, that might be brought to bear in treating the modal version of Evans's argument, cannot be correctly wielded in the many-valued case. This is van Inwagen's objection and one of Parsons's objections to Evans's argument. Parsons remarks that Evans's argument shows us "that we cannot extend property abstraction to formulas containing indeterminacy operators." I have just tried to show that such an objection is groundless, and not available to the believer in a many-valued account of vagueness.

Parsons and Broome have made detailed suggestions about how we should interpret → (and Leibniz's law) in their favorite three-valued logics. I find myself less than enthralled with their suggestions, but rather than discuss their proposals, I shall content myself with providing a somewhat different argument against vagueness-in-reality. The overall structure of my position is: a many-valued logic is the only legitimate logic that a vagueness-in-nature theorist can use. Such a theorist is required to be able to "speak about" vagueness and therefore must have at his disposal some device such as our J operators. Further, whatever principles the theorist uses must agree with classical logic on the classical values. Finally, there are some principles governing the use of J operators, Leibniz's law, the reflexivity of identity, and the interaction of J operators with quantifiers; I argue these principles to be correct, and anyway most "vagueness-in-nature" theorists believe them. I shall first lay out all these principles and then provide an argument that shows that no three-valued logic obeying the principles can admit "vagueness-in-reality." Not every three-valued logic is susceptible to my argument. The ones that are have the following properties: they have the resources to "talk about

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

vagueness" (either directly by having J operators or by having other operators with which the J operators can be defined), and they have sufficient connectives in the language to allow us to define certain other connectives, here called &, $\vee$, $\rightarrow$, $\sim$, $\leftrightarrow$, which in turn operate as specified below and which operate classically on the classical values (1 and 3).[12] But this is a very large group of three-valued logics; and I would say that any logic that did not obey these restrictions should not be taken seriously as an account of our reasoning about vagueness. Then I shall show how a simple generalization of certain of the J principles will prove that no finitely many-valued logic obeying certain simple restrictions can accomodate "vagueness-in-nature." (That is, my basic argument does not turn on any such concept as "higher-order vagueness.")

We start with some principles governing J operators. The first principle is "uniqueness of semantic value" (USV), saying that every formula takes exactly one of the three semantic values:

$$\text{(USV)}\quad (J_1\Phi \vee J_2\Phi \vee J_3\Phi) \,\&\, \sim(J_1\Phi \,\&\, J_2\Phi) \,\&\, \sim(J_1\Phi \,\&\, J_3\Phi) \,\&\, \sim(J_2\Phi \,\&\, J_3\Phi)$$

Second, we interpret the semantic value 1 as "completely true" and therefore insist that, if $J_1\Phi$ is true, then ϕ cannot be "completely false," i.e., that

$$(J_1)\qquad J_1\phi \rightarrow \phi$$

is an axiom. Note that (J_1) does not prejudge whether ϕ is or is not "completely true," only that, if it is "completely true," then it is "true". Third, the $\leftrightarrow$ connective obeys this rule of inference:

(E) if $(\phi \leftrightarrow \psi)$ is semantically valid, then so is $(J_i\phi \leftrightarrow J_i\psi)$, for any $1 \leq i \leq 3$

In fact, (E) is no restriction on the logic at all. It amounts to assuming that, if a biconditional formula is provable, then any bicondi-

[12] Just as in the case with J operators, where it was not required that the language have them as primitives but only that they be definable, here too we do not insist that the language have these sentential connectives—only that they be definable with the resources that the language does have. By 'operate classically on classical values' I of course mean that '&' operates like the classical 'and' on the classical values, that '$\vee$' operate like the classical 'or' on the classical values, and so forth.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tional formula expressing an instance of the diagonal is also provable. (A biconditional formula "expresses an instance of the diagonal" iff both sides have the same truth value. Any logic that makes the $\leftrightarrow$ true when both sides have the same truth value obeys this constraint.)

With regard to identity, the logic is to have two properties: first,

(refl′) $$J_1\alpha = \alpha$$

is an axiom; and, second, it must have a version of Leibniz's law in it which entails all the instances that this one entails: that, if a and b are identical, then, for each property, it cannot happen that one has it and the other lacks it, that is,

(LL) $$a = b \leftrightarrow (\forall F)(Fa \leftrightarrow Fb)$$

And, finally, with regard to the interaction between quantifiers and J operators, the only principle to be used is this:

(J_2-$\forall$) $$J_2(\forall x)Fx \rightarrow \sim(\exists x)J_3Fx$$

which can be read, roughly, as "if it is true to degree 2 that everything is F, then there cannot be anything of which it is completely false that it is F." Whatever the difficulties are in giving an explicit and complete statement of the interrelationships among the J operators and the quantifiers, this much at least seems clear: if there were an object for which it was definitely false that it was F, then it cannot be vague/indeterminate/undecided whether everything is F—on the contrary, it would be completely false that everything is F.[13]

One lemma needs to be mentioned, which follows from (USV) and the fact that the J operators are two-valued. Recall that, if any J operator is applied to any sentence, its value will be either 1 or 3. Therefore, putting a J_2 in front of a sentence that already has a main J operator will yield the value 3. More generally, if each sentential part of a formula ϕ is already in the scope of some J operator, prefixing a J_2 operator to ϕ will yield the value 3; that is, its negation

[13] The use to which I put this principle involves the second-order quantifier $(\forall F)$; but (J_2-$\forall$) applies to such quantifiers just as much as it does to first-order ones. Whatever reasons one thinks it holds in the first-order case can be carried over to the second-order case. The principle is independent of the order of the quantifier.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

will yield the value 1 (since negation operates classically on the classical values):

(Lemma) $\sim J_2\Phi$ (equivalently, $J_3\Phi$), if all sentences in ϕ are already in the scope of some J operator

Now let us look at the argument. Suppose an object a is vague. According to the vagueness-in-reality theorist, this implies that there will be a true vague identity statement, that is, a statement of the form $J_2\,(a = b)$.

a. $J_2\,(a = b)$	Assumption
b. $a = b \leftrightarrow (\forall F)(Fa \leftrightarrow Fb)$	(LL)
c. $J_2\,(a = b) \leftrightarrow J_2\,(\forall F)(Fa \leftrightarrow Fb)$	from (b) and (E)
d. $J_2\,(\forall F)(Fa \leftrightarrow Fb)$	from (a), (b) $\leftrightarrow$ elimination
e. $\sim(\exists F)\,J_3\,(Fa \leftrightarrow Fb)$	from (d) and (J_2-$\forall$)
f. $(\forall F)[J_1(Fa \leftrightarrow Fb) \vee J_2(Fa \leftrightarrow Fb)]$	from (e) and (USV)
g. $J_1\,(J_1\,(a = a) \leftrightarrow J_1\,(a = b)) \vee J_2\,(J_1\,(a = a) \leftrightarrow J_1\,(a = b))$	from (f), instantiate $(\forall F)$ to $(\lambda x)\,J_1\,(a = x)$ and λ convert
h. $J_1\,(J_1\,(a = a) \leftrightarrow J_1\,(a = b))$	from (g), (lemma), disj.syll.
i. $J_1\,(a = a) \leftrightarrow J_1\,(a = b)$	from (h) and (J_1)
j. $J_1\,(a = b)$	from (i) and (refl′), $\leftrightarrow$ elim.
k. $\sim J_2\,(a = b)$	from (j) and (USV)

Now, each step is legitimate here. All propositional inferences ($\leftrightarrow$ elimination, disjunctive syllogism) were done on classical values and all other steps were applications of the principles listed, in which I include the legitimacy of using the indicated instance of the quantifier in step (g), as argued earlier. Although "ϕ and $\sim\phi$" is not in general an impossible state of affairs in a three-valued logic, here (k) and (a) really are jointly impossible, since they both have classical values.

A structurally identical argument can be made within any finitely many-valued logic (so long as it has only one designated value). The only differences are that certain of the background principles need to be altered: given an n-valued logic (n being "most false"), we have

(USV′) $(J_1\Phi \vee J_2\Phi \vee \cdot\ \cdot\ \cdot \vee J_n\Phi)\ \&\ \sim(J_i\Phi\ \&\ J_k\Phi)$, for $i \neq k$

(E′) if $(\phi \leftrightarrow \psi)$ is semantically valid, then so $(J_i\phi \leftrightarrow J_i\psi)$, for any $1 \leq i \leq n$

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(J_i-$\forall'$) $J_i(\forall x)Fx \rightarrow \sim(\exists x)J_n Fx$, for any $i \neq n$

(lemma′) $\sim J_i\phi$ (for any i other than 1 and n), if all sentences in ϕ are already in the scope of some J operator

which are all clearly as justified in an n-valued logic as their correspondents were in the three-valued case. The principles

(J_1) $J_1\phi \rightarrow \phi$

(refl′) $J_1\alpha = \alpha$

(LL) $a = b \leftrightarrow (\forall F)(Fa \leftrightarrow Fb)$

remain the same. The argument now is: suppose an object a is vague, then the vagueness-in-reality theorist is committed to $J_i\,(a = b)$ for some $1 < i < n$.

a. $J_i\,(a = b)$	Assumption
b. $a = b \leftrightarrow (\forall F)(Fa \leftrightarrow Fb)$	(LL)
c. $J_i\,(a = b) \leftrightarrow J_i\,(\forall F)(Fa \leftrightarrow Fb)$	from (b) and (E′)
d. $J_i\,(\forall F)(Fa \leftrightarrow Fb)$	from (a), (b) $\leftrightarrow$ elimination
e. $\sim(\exists F)J_n\,(Fa \leftrightarrow Fb)$	from (d) and (J_i-$\forall'$)
f. $(\forall F)[J_1\,(Fa \leftrightarrow Fb) \vee \ldots \vee J_{n-1}\,(Fa \leftrightarrow Fb)]$	from (e) and (USV′)
g. $J_1(J_1\,(a = a) \leftrightarrow J_1\,(a = b)) \vee \ldots \vee J_{n-1}\,(J_1\,(a = a) \leftrightarrow J_1\,(a = b))$	from (f), instantiate $(\forall F)$ to $(\lambda x)J_1\,(a = x)$ and λ convert
h. $J_1\,(J_1\,(a = a) \leftrightarrow J_1\,(a = b))$	from (g) and (lemma′)
i. $J_1\,(a = a) \leftrightarrow J_1\,(a = b)$	from (h) and (J_1)
j. $J_1\,(a = b)$	from (i) and (refl′)
k. $\sim J_i\,(a = b)$	from (j) and (USV′)

Finally, the argument also is valid in any infinite-valued logic in which we can formulate it (and which has principles similar to those listed above). The restriction about being able to formulate it amounts to requiring at most a denumerable number of truth values, or else the language will not have the ability to form the required number of J operators. All the principles listed above seem true in any plausible infinite-valued logic, including the (J_i-$\forall$) principle. [Often in infinite-valued logics the semantic value of $(\exists x)Fx$ is defined as the limit of the values of $F\alpha$, for all α; and this means there need be no particular α for which $F\alpha$ takes that value. This, in turn, means that there is no easy way to state a general relationship

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

between J formulas with embedded quantifiers and those where the quantifier is outside the scope of the J operator. But, since we require only an analogue of $(J_i\text{-}\forall)$, such difficulties will not arise for our argument.[14]]

These arguments are similar to Evans's original one, with the exception of making it explicit that we are operating in a many-valued environment and making explicit precisely what principles are involved. The conclusion is: no standard logic, not even a standard many-valued logic, can admit vague objects, if that is taken (as our theorists wish) to include the possibility of making vague identity statements, under pain of explicit, two-valued contradiction. Parsons has objected to this conclusion on the grounds that "it proves too much." If this argument is correct, he says, then a similar argument "can be used to refute the existence of indeterminacy in truth value of ordinary claims." I agree that the conclusion follows (under the plausible requirements laid out above), but draw a different conclusion than the one implicit in Parsons. Rather than show the impossibility of vagueness, or indeterminateness, I think it shows the incorrectness of analyzing these concepts by means of a many-valued logic. Instead, vagueness ought to be viewed as a semantic notion and investigated by means of different evaluation techniques laid atop a classical logic.[15] But this is just to deny that there is any such thing as vagueness-in-reality.

FRANCIS JEFFRY PELLETIER

University of Alberta

[14] For a different set of objections to such "fuzzy logics," see C. Morgan and my "Some Notes on Fuzzy Logic," *Linguistics and Philosophy*, I, 1 (1977): 87–121.

[15] About the penultimate sentence of the text, Sylvan wishes me to emphasize that a relevance logic might be used instead of a classical logic. But we agree on the point that vagueness is a semantic notion to be laid atop the logic.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

BOOK REVIEWS

The Structure of Emotions. ROBERT M. GORDON. New York: Cambridge University Press, 1987. xiv, 161 pp. $29.95.

The Rationality of Emotions. RONALD DE SOUSA. Cambridge: MIT Press, 1987. xix, 373 pp. $25.00.

On a familiar model for explaining action, belief and desire play an essential and irreducible role, simultaneously specifying what caused the agent to do what she did and identifying her reasons. Emotion surely also enters into the explanation of action; but just how it does is not clear. What is the nature of the emotions? What relation do they have to beliefs and desires? In particular, is the relation like that between beliefs, desires, and actions? In what ways are emotions subject to rational appraisal? And, finally, do the emotions suggest a way in which the belief-desire model must be supplemented?

Both Ronald de Sousa and Robert Gordon address these questions in one way or another. Since Gordon addresses them more explicitly, his book provides the better point of departure. Gordon claims that a number of emotions are caused by, and identified in terms of, beliefs and pro/con attitudes. In this respect, the explanation of emotions resembles the explanation of actions. Yet the emotions are not reducible to beliefs and desires, nor any combination of them. Emotions fill in "serious gaps in our attempts to explain human behavior in terms of beliefs and desires or their scientific counterparts" (7 & 8), and thus are a needed supplement to the reasons-explanation model.

Two examples carry much of the burden of Gordon's argument. In one example, farmer *A* and farmer *B* both believe there is a 50-50 chance of rain, and both have the same desire for the flourishing of their crops. Yet *A* does and *B* does not prepare to irrigate. Thus, *A* acts as if he believed it would not rain, and *B* as if he believed it would. But, by hypothesis, neither believes that it will rain nor that it will not; *A* and *B* are equally uncertain. The story is not implausible, Gordon maintains. "Yet if motivational differences could be accounted for only by differences in beliefs or desires, the story would be *impossible*" (8).

Gordon suggests that the extra ingredient is supplied, if we know that farmer *A* is afraid it will not rain, whereas farmer *B* is hopeful it will. What fear adds to the story is a certain "motivational effect." "To describe the effect in very broad terms, one is motivated to take things into one's own hands: to increase the scope of one's control

0022-362X/89/8609/493–504 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

over the world, to decrease the scope of one's helplessness" (67). What Gordon calls "the functional resemblance" between fearing that p and believing that p consists in the fact that, in attempting to avoid vulnerability, one acts as if one believed that p were true.

In the second example, a woman burns some manuscripts of her husband's lover, not in the belief that doing so will somehow remove the other woman from her life, nor from a desire to get even. To exclude these possibilities, Gordon supposes that, prior to the wife's action, both her husband and his lover have died and that the wife has known of their deaths. Here the missing element in the belief-desire explanation is supplied by the emotion of anger, which Gordon says is a response to wish frustration. Fear motivates one to increase one's control in situations in which one is uncertain about something that bears on one's wishes; anger motivates one to harm the person one believes responsible for frustrating one's wishes (in this case the wish that the husband had not had the affair). It is such motivational effects, presumably a different one for each emotion type, that the emotions peculiarly supply in explanations of behavior.

The structure of Gordon's explanation of the case of the two farmers depends on the thesis that propositional fear necessarily swings the decision one way rather than the other, that is, toward reducing vulnerability and increasing the scope of control. But is it clear that simply knowing A fears it will not rain tells us he will be motivated to irrigate? Might he not be paralyzed by his fear, or motivated to deny it, acting in either case as someone might who believed it would rain? If fear could push A either way, then adverting to fear fails to explain his action, just as beliefs and desires were said to fail.

In any case, the natural way to interpret the fact that A is afraid it will rain while B is not is that A's desire to avoid the consequences of crop failure is greater than B's, perhaps because of other beliefs and desires or perhaps because of some third factor. Either way, appeal to the emotions does not close a gap in the belief-desire explanation, but makes it a bit more subtle by introducing a consideration of strength of desire. It is unlikely that anyone has ever claimed that actions can be predicted on the basis simply of knowing what beliefs and desires an agent has while knowing nothing of their relative strengths.

Perhaps Gordon's point is rather that, even when the subjective probability of belief and the relative strength of desire are the same for A and B, they may act differently, and that it is emotion that makes for the difference. If this were how emotions entered into the explanation of action, they would not be factors bearing on the

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

beliefs and desires but on the action directly. On this interpretation, it would appear that emotions are not reasons the person has but forces acting on him. The idea of emotions as basically irrational or nonrational forces has a long history, but this seems to be one of the ideas Gordon wants to disarm. In any case, the fact that emotions may lead to irrational behavior does not itself show that the emotions explain actions in a way beliefs and desires cannot; beliefs and desires sometimes explain irrational actions.

If emotions may be counted among the rational causes of action, yet different from beliefs and desires in the way they explain actions, what is the difference? Gordon says that emotions explain motivation in a way that beliefs and desires do not. But, since on any plausible interpretation of the belief-desire model, 'desire' must be understood very broadly, as Gordon recognizes, this leaves him with the problem of showing how "motivated to increase one's control," for example, differs from a desire (broadly conceived) to increase the scope of one's control; or how being motivated to harm someone differs from a desire to harm her.

Gordon thinks there is a basic difference. He claims that desiring and wanting are not the requisite pro-attitudes in the case of the emotions. When *S* is angry or delighted about the fact that *p*, *S* already knows that *p*. "This leaves S with no possibility of *instrumental or aversive* action. In such a case the notions of wanting and desiring are usually thought inapplicable" (30).

This distinction between wishing and desiring is crucial to Gordon's analysis of the difference between the ways in which emotions and actions are justified. "What emotions *share* with actions is, most importantly, an origin in propositional attitudes rather than in 'brute' causes exclusively" (111). It is this propositional component which distinguishes the emotion of fear, peculiar to human beings, from the biological state of fear which human beings can share with other animals; or the emotion of anger from a purely neurological or hormonal disturbance. Sue is angry with Ellen because Sue believes that Ellen has made off with Sam, and Sue cares about that because she wishes Sam were hers alone. These propositional components are the reasons Sue is angry. So far the relation between beliefs, desires, and actions, on the one hand, and beliefs, wishes, and emotions, on the other, is the same. Yet desiring engages with means-end beliefs about how to satisfy the desire and so leads to intentional action; wishing does not engage with such beliefs (Richard Wollheim[1]

[1] *The Thread of Life* (New York: Cambridge, 1984).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

has made a similar point about something he calls phantasy) and leads instead to emotion. Since one's reasons for acting include an instrumental belief that so acting is the way to accomplish one's end, one's desires and beliefs "argue for," "justify," "say something in favor of," performing that action (122). Yet the attitudes and beliefs that underlie an emotion do not argue for, say something in favor of, or justify one's having that emotion. This difference explains the fact that we do not reason with someone about his emotional response in the way we reason with him about the rightness or wrongness of his action.

But the reasons for an emotion do at least sometimes justify it in the sense of showing it to be broadly reasonable, or appropriate. Say I hold that stealing is shameful, and I believe I have just stolen something. It would seem this both explains and justifies my shame. And, sometimes, believing that someone's response to a situation would be more appropriate were he to have an emotion he does not have, we try to encourage that response in him by pointing out facts he has not noticed, or has not acknowledged, or the significance of which he has denied, by reminding him of certain of his desires or values.

Furthermore, is it the case that in Gordon's examples of wishing there is no instrumental or means-end belief? Does farmer *A* not believe that preparing to irrigate is a means of satisfying his wish that his crops be watered *come rain or come shine,* or his wish to reduce his anxiety? Does the angry wife not believe, however irrationally, that burning the dead woman's manuscripts is a means of harming her? An irrational belief is still a belief. If we do need to invoke such instrumental beliefs as these to explain the behaviors in question, then the difference between wish and desire is not the logical difference Gordon claims it is, or at least not for the reason he gives.

The pro-attitude in emotion cannot be desire, Gordon says, nor is it judgment or evaluation. For "it is at least arguable that evaluative beliefs do not necessarily involve the corresponding attitudes: that one may believe something to be 'conventionally' good (bad) . . . while being quite indifferent to it oneself . . ." (30). Of course, the latter is possible. But that someone might truly believe something to be good, not just conventionally good or good in the eyes of others, yet be indifferent to that good, is an idea many have found incoherent. If this is Gordon's basis for distinguishing emotions from evaluative judgments, the distinction is on shaky ground.

What is helpful about Gordon's book is the variety of ways in which it shows how to accommodate emotions to the reasons-explanation model. For example, so-called forward-looking emotions like fear

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

have seemed to pose an obstacle to the view that the cause of an emotion is also its reason, since the future war that may be the "object" of or reason for someone's fear cannot be the cause of her present emotional state. To handle these cases and to do the work of talk about the object of an emotion, Gordon introduces a useful distinction between what he calls *factive* and *epistemic* emotions. Emotions depend on beliefs of some sort, and typically on some positive or negative attitude. For the factive emotions, " 'S emotes (e.g. is amazed, is angry, is delighted) that *p*' is true only if it is true that *p* and, further, S knows that *p*" (43). For the epistemic emotions, " 'S emotes (e.g. is afraid, is hopeful, is worried) that *p*' is true only if S is not certain that *p* (and therefore cannot be said to know that *p*)" (43). Unlike anger and disappointment, fear essentially involves uncertainty, and it is this epistemic state which provides the causal belief condition. The distinction between factive and epistemic emotions also allows Gordon to discriminate the different reasons in the form of both belief and attitudinal conditions that identify a number of emotions.

As is clear, Gordon's account of the emotions uses the reasons-explanation model as a basis. He does not think he needs to raise any fundamental questions about the concepts of mind or rationality. In this respect, de Sousa's is the more radical view. Like Gordon, he holds that the emotions are mental phenomena that are not reducible to beliefs and desires. But in support of these claims de Sousa offers a new analysis of what constitutes the mental, an analysis that attempts to place emotion in an evolutionary scheme. Gordon argues that emotion enters into the explanation of action in a special way. De Sousa more ambitiously contends that emotion is a necessary condition for the exercise of human reason.

In de Sousa's view, mentality or intentionality comes in different grades or degrees. At the low end of the intentional spectrum is quasi or nonmental intentionality: stimuli are translated into stereotypical sequences of "T-instinctual" behavior by automatic mechanisms. "An intermediate grade involves representations, bearing semantic and logical relations to a network of other representations" (78). At the high end of the spectrum is full mentality or intentionality, the specific characteristic of which is the capacity for "singular reference."

The reason de Sousa gives for calling T-instinctual behavior a grade of intentionality is that T-instinctual sequences involve feedback, the testing and evaluation of which must already have some characteristics in common with representation. These characteristics are "generality" and "intentional inexistence," namely, that "the

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

condition 'represented' in the testing mechanism may never actually occur" (97). It is not clear from de Sousa's account that these two representational characteristics refer to anything other than a creature's capacity for a discriminatory response to the environment. If they do not, then thermostats and plants, as well as worms, are quasi-intentional mechanisms, and the concept of intentionality has lost touch with the mental. The sense in which a thermostat may be said to represent something would seem to have little to do with the sense in which we can think about ("represent," in de Sousa's construal) golden mountains.

A more fundamental question about this scheme is the place it gives to singular reference, which de Sousa says is possible only for a creature with the capacity "to intend a logical distinction even between qualitatively indiscriminable individuals" (100). Singular reference is the distinguishing mark of something de Sousa calls nonfungible identifications. An example of what he has in mind, he tells us, is Alcmene's refusal to settle for a substitute she knows to be such even though she cannot tell him apart from her beloved Amphitryon. It is true that we want to distinguish normal adult love from Fido's love for its master, or an infant's love for its mother, and that this distinction contains some notion about how the lover conceives of the love object. (De Sousa refers in this context to "transitional objects" [99], the psychoanalytic idea of a pre-linguistic stage in which a child's earliest conception of another person does not distinguish her clearly from itself. Presumably de Sousa's point is that such a stage is transitional to the capacity for singular reference. But he does not say what, if anything, the psychoanalytic idea adds to our understanding of this capacity. This is an example of the sort of interesting connections he often makes, but also of his failure to develop them.) The capacity for singular reference may capture something important about human love. But is this capacity a plausible criterion for full intentionality? If one has the concept of a person, or an animal, one must know what it is to be one person or animal rather than another. Singular reference is part of a scheme that includes predication and the concept of identity, so there seems to be no reason to give one element more weight than another in an analysis of the mental. Perhaps de Sousa's idea is that the capacity for love in the sense that singular reference articulates is a necessary condition for mind. That would be an interesting claim; and, if true, it would certainly support his thesis that emotion is an essential ingredient in rationality. But the claim would have to be argued.

De Sousa tells us that the notions of intentional spectrum and singular reference supply necessary reinforcement to Freud's argu-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ments about the unconscious, for Freud needs a concept, mid-way between need and desire, to show how desire itself is shaped. Instinct plays this role for Freud, but the difficulty with his account is that instinct, though unconscious, is still supposed to be a psychological concept. How the concept of instinct can be justified, de Sousa says, "is the problem of the justification for the concept of unconscious mind" (94).

This analysis of the role of instinct in Freud's thought seems right. But de Sousa confusingly elides the instinctual with the unconscious. The concepts are not interchangeable for Freud. Sometimes he uses the unconscious as a name for fully mental yet not conscious beliefs and desires, identified as such by their place in a holistic network of other beliefs and desires. This is one way of identifying what it is the conscious and unconscious have in common. On this interpretation, the justification for the concept of the unconscious would be the obvious one, that unconscious beliefs and desires are often needed to explain human behavior.

Why, in any case, is consciousness so important that we have to justify calling what is unconscious mental? It required a justification for Freud, because he had inherited a Cartesian tradition and it was within its assumptions that he tried to state his discoveries. Perhaps de Sousa also believes that consciousness is a constitutive feature of the fully mental. This would explain his correlating unconscious mental states with a lesser degree of intentionality. But, in this case, the idea that there might be unconscious beliefs and desires becomes problematic.

Even if we grant the intentional spectrum, de Sousa's own answer to the question of how desire is shaped is confusing. The concept Freud needs, de Sousa says, the concept that is to replace and fill out Freud's instinct, is "H-instinct," which, unlike T-instinct, produces different patterns of goal-oriented behavior. H-instinct is prior to full intentionality. Yet de Sousa says that T-instinct becomes H-instinct by the addition of "some crucial factor," which he goes on to identify as the capacity for singular reference (97). In this case, H-instinct and the unconscious are fully intentional, and H-instinct may shape desire in the ways that beliefs and desires themselves do. H-instinct and the emotions are classed together as factors preceding full intentionality, factors that "determine motivation" (84). I take this to mean that they are factors that explain why we have the desires we do. But, if they are to play this role, we need to know just how emotion and H-instinct differ from desire.

Since it is plain that de Sousa means to distinguish H-instinct from full intentionality, one might dismiss this unclarity as an oversight

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

but for the fact that in other ways he leaves us in the dark about a critical question, namely, the relation between emotions and propositional attitudes. His distinction between fungible and nonfungible identifications, or identifications that are not made in terms of singular reference and those which are, applies also to the emotions. Does this imply that nonfungible emotions, the ones that presuppose full mentality, have a propositional component? De Sousa does not say. Furthermore, fungible emotions are presumably pre-propositional. But his description of the grade at which emotions enter the intentional spectrum is insufficiently elaborated for us to understand the relation between the representations said to characterize this grade, concept formation, and propositional attitudes. De Sousa's fuzziness on these issues clouds his claims that emotions are irreducible to beliefs and desires, and that emotions are subject to rational appraisal in a way peculiar to them.

Where Gordon's argument depends on distinguishing desire from wish, de Sousa's depends on distinguishing desire from want. De Sousa holds that it is want, not desire, that enters into ordinary reasons-explanations, and that desire, as the pro-attitude at work in the emotions, cannot enter directly into the reasons-explanation model. (Nor, presumably, can wish, since de Sousa holds that it, too, *is* an emotion [169]). This is not because desire and emotion are in any sense irrational, but because they ground reason as we ordinarily understand it. De Sousa writes:

> Rationality never prescribes, but only constrains, by proscribing inconsistency and distinctions without a difference . . . (163).

> What is needed for rational action is not desire but something weaker. Call it *functional wanting*, which needs to satisfy only the conditions of quasi intentionality. If a computer is programmed to pursue certain goals and to find means to those goals, it is thereby given functional wants (167).

There are two paradigms of rationality, the argument continues, cognitive and strategic. The first is the relevant paradigm for beliefs, the second for wants. A belief is cognitively rational if arrived at in a way likely to make the belief true. A want is strategically rational to the extent that it is consistent with the agent's other wants. But human agents not only calculate the means to predetermined goals; they also choose the goals, and not only on the basis of their consistency with other goals. What is needed to account for such choice, de Sousa concludes, is desire, which he takes "as a paradigm of emotion" (166). Therefore, if the ways in which one goes about the

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

choice of a goal can be said to be rational or irrational, a third paradigm of rationality, the one appropriate to emotion/desire, is needed.

The key to this argument, as becomes clear in the chapter on the rationality of emotion (171–203), is the assimilation not only of desire but also of valuing to emotion. It is valuing which is needed for the full exercise of reason and which neither cognitive nor strategic rationality can assess. De Sousa supports his claim that emotions are essential to rationality, though they are not among its mechanisms, by denaturing the concept of reason, by construing reason so as to exclude desire, and by making the emotions do the explanatory work of both desire and valuing. Having in this way stripped reason of motivating energy, he can also claim that emotion is needed "to break the tie when reason is stuck" (16).

The criteria of rationality appropriate to the emotions can be educed from their resemblance to perceptions, de Sousa claims. Emotions are "apprehensions of real properties in the world" (201), value or "axiological" properties. These properties then provide an objective standard of correctness for all the emotions, and a sense in which the emotions themselves, when correct, can be said to be objective. This is de Sousa's answer to what he poses in the first chapter as "the Euthyphro question": Is something lovable simply in virtue of being loved, or is it loved because of some properties intrinsic to it?

Emotions resemble beliefs and desires in having conditions of success, formal objects which delimit the conditions of their intelligibility. "The notion of formal object applies to any state having a content assessable for correctness: it is then, by definition, the *standard of correctness* for that state" (122). Truth is the formal object of all beliefs, and good the formal object of all wants or desires. The formal object of an emotion is the property the cause of the emotion must have if the emotion is correct. If one is appropriately ashamed or afraid, the condition or object in the world which is causing one to have that emotion must be shameful or fearful. More particularly, the formal object of an emotion is "a second-order property that . . . is supervenient on some other property or properties" (122). For example, something is fearful in virtue of being dangerous. (It is harder to specify the properties appropriate to other emotions. I come to this point in discussing paradigm scenarios below.) Thus, unlike beliefs and desires, the emotions do not form a single class all of whose members have the same formal object. This is one of de Sousa's several arguments for the contention that emotions cannot be reduced to beliefs and desires.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

One longs in this section (107–137) for specification of the objects of the emotions in terms of propositional attitudes. It is not that de Sousa ignores such specification altogether; but he distinguishes so many other kinds of objects that the significance of any one is lost. In any case, the notion of formal object does not help clarify the criteria of rationality appropriate to the emotions. First, it is not obvious that each emotion type has its own formal object. What, for example, is the formal object of jealousy, of pride, of expectancy? The plausibility of the claim seems to derive mainly from a grammatical fact that, for some of the emotions, there is a parallel pair of predicates to describe the emotion and the object causing it. (Love the lovable, fear the fearful, etc.) But many emotions lack an obvious formal object. Second, if emotions are objective in the way de Sousa claims, it seems that truth can serve as the formal object for them all, in which case the emotions can be assessed by the same criteria of cognitive rationality appropriate to beliefs. Finally, de Sousa's claim about the objectivity of the emotions depends on a prior and unargued thesis about the objectivity of value judgments, which I take to be the question whether they can be true or false.

Exploring the relations between emotions and axiological properties is complicated by the fact that every person's emotional responses are a function of something de Sousa calls paradigm scenarios, "little dramas in which our natural capacities for emotional response were first enlisted" (48). The concept of paradigm scenarios is meant to explain the ways in which, despite their objectivity, emotions are relative to the individual. It is meant also to tell us something about how the mental emerges from the nonmental. Emotions are mental states that are constructed out of innate biological dispositions. A model for such a process, de Sousa suggests, is the way in which the smiling response of the infant, a response which is at first purely physiological, comes to "mean" something through the infant's social interactions with others (182). Similarly, a child "is genetically programmed to respond in specific ways to the situational components of some paradigm scenarios. . . . An essential part of education consists in identifying these responses, giving the child a name for them in the context of the scenario, and thus teaching it that it is experiencing a particular emotion" (183). De Sousa also claims that paradigm scenarios "define the very character of our emotions" (183). It is not clear whether he thinks of these as different claims, though they are. Paradigm scenarios are meant to provide "the first step in search of a *semantics* of emotion" (181). De Sousa does not develop this idea, and one wonders if he could.

Emotional irrationality, he claims, is in part a function of a per-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

son's mistakenly perceiving a new situation in terms of an inappropriate paradigm scenario. Beyond that, our emotions should be appropriate, not only to our own, but also to the widest possible range of paradigm scenarios. He calls this "the principle of emotional continence" (187). De Sousa's intuitions here seem right: on the one hand, emotions are rooted in a more direct way than other mental states in an individual's early history, and an account of both character and emotional irrationality will have to reckon with these histories; on the other hand, it is in our emotional responses that we most finely acknowledge certain truths. So it is disappointing that these intuitions are not better served. The principle of emotional continence remains a dark notion, a defense of which would in any case presumably require a better explication of the thesis about axiological properties. As for the concept of paradigm scenarios in general, nothing in the social process, as de Sousa describes it, through which the emotions are constructed is peculiar to them. Perceptions, beliefs, and desires, at least on some views of the mind, are constructed in much the same way. We learn to identify things in the world through a process that essentially involves other people, and this process might be said to define the very character of our thoughts and so of our mental life. Further, one wonders what the situational components are to which the child is genetically programmed to respond in learning an emotional vocabulary. The only example de Sousa gives is the development of the smiling response, and a smile is not an emotion. The concept of paradigm scenario does not really clarify, as it is meant to, the relation between emotions and the properties of objects in virtue of which emotions can be objective.

De Sousa says that the emotions do not figure as reasons in patterns of explanation, but are "a species of determinate patterns of salience among objects of attention, lines of inquiry and inferential strategies" (196). This may be taken to be another argument in support of the irreducibility of emotions to beliefs and desires. It is true that when one loves, or is angry, or fearful, one looks at things differently. Emotion may give a new color to beliefs and desires one already has, or prompt new attitudes altogether. But, if this is what de Sousa means in saying (in the same passage) that emotions "set the agenda for beliefs and desires," then there is no reason to think that emotions are not themselves explicable in terms of such propositional attitudes, at least if one interprets desiring, as is usually the case, to include valuing.

Again, I think de Sousa is onto something. It seems right that we can attribute emotions to infants earlier than we can attribute propositional attitudes to them, and that the patterns for emotional

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

responses are learned early in infancy, specifically through interpersonal relations. It may also be that emotions are the mental precursors of beliefs and desires and that, in de Sousa's provocative phrase, they set the agenda for them. These are the conclusions de Sousa has tried to set up in his earlier discussions of an intentional spectrum. But the claims that, before there are any propositional attitudes, there are mental states in the form of anger, fear, and so on, and that such mental states partially determine what our first beliefs and desires will be, are the very claims that need to be better clarified and argued than I find them to be anywhere in the book.

In passing, I remark that, of all the emotions, love is the most difficult to analyze as a propositional attitude. There is no set of propositions which will capture the content of Alcmene's love for Amphitryon. I do not know what this reveals either about love or about emotion more generally. But it cannot be incidental to other differences between Gordon and de Sousa that one (Gordon) does not discuss love, while the other places it at the center of his theory.

I have neglected much that is of interest in these books in order to concentrate on questions central to both: the relation between emotions, beliefs, and desires; and the role of emotion in the exercise of reason. De Sousa makes many claims for emotion; and some of these claims are on the face of it paradoxical: desire is paradigmatic of the emotions, yet emotions are not desires; emotions are apprehensions of real properties in the world, yet emotions are not beliefs; emotions are valuings, yet not judgments. Nevertheless, de Sousa has opened paths which lead from the emotions to key issues in the philosophy of mind and ethics, paths which other philosophers will want to explore.

MARCIA CAVELL

University of California/Berkeley

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVI, NO. 10, OCTOBER 1989

THE METAPHYSICAL CONSTRUCTION OF VALUE*

IN his third Carus lecture and the *Reply,* H. P. Grice[1] argues for the legitimate attribution of value—value that is not relative and not reducible. The lecture divides into two important discussions: the first on teleology and finality, the second on what he calls metaphysical construction routines. What is of special interest is not that Grice thinks value is constructed by us, but that he thinks this brings objectivity with it. Grice was at one with Nelson Goodman's "skeptical, analytic and constructionalist orientation" and the "rigorous constraints" on construction which this imposes; but he would deny that constructivism is necessarily coupled with "radical relativism . . . that eventuates in something akin to irrealism."[2]

I shall try to explain how construction routines are supposed to work, and raise problems only at the end. I suggest that these metaphysical operations are best understood, not as invented procedures, but as natural dispositions to take the world to be a certain way; under certain conditions, they bring reason with them. The specified conditions legitimate our natural tendencies; we not only take things to be thus and so and act as if they are, but we are justified. We can move from the fictive and the "as if" to what there *is* in the world.

Two metaphysical routines are essential to the argumentation for value: *Humean projection* and *metaphysical transubstantiation.* The first is meant to explain how it is that our judgments of value

* To be presented in an APA symposium on the Thought of Grice, December 28. Richard E. Grandy will be co-symposiast, and Robert Stalnaker and Richard Warner will comment; see this JOURNAL, this issue, 514–525, 526/7, 528/9, respectively, for their comments.

[1] "The Conception of Value," Carus Lectures, 1983, to be published by Oxford University Press. "Reply to Richards," in *Philosophical Grounds of Rationality: Intentions, Categories, Ends,* Richard E. Grandy and R. Warner, eds. (New York: Oxford, 1986), pp. 45–108, cited as *Reply.*

[2] *Ways of Worldmaking* (Indianapolis: Hackett, 1978), pp. 1 and x.

0022-362X/89/8610/505–513 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

both ascribe a property to something in the world and motivate us; it constructs or delivers a concept such that this hybridity is accounted for. Grice was keenly aware that supervenience, far from solving this problem, is merely a restatement of it.

The second routine permits us to say that there is a kind of creature whose essence, nature, or function is to evaluate. Grice hoped to show that the notion of value is indispensable. In the lectures, he said: "One would first set up a specification of a series of increasingly complex creatures; and one would then exhibit the notion of value as entering essentially into the theory attending the last and most complex type of creature appearing in that series." At each stage of this development, one must offer theoretical justification for the next stage. One could then, he thought, claim a legitimate metaphysical demand to construct such a creature.

This is not to say that one has thereby shown that some things are of value. The *notion* of value enters essentially because persons evaluate and demand justification, ultimately in terms of nonrelative value. Nonetheless, it is an important anchor for our conception of value that it is part of the fiber of the kind of thing we are and not easily expunged.

Moreover, the notion of a creature whose essence is to evaluate permits the application of 'good of its kind' to persons insofar as they do their job of acting in line with justifiable beliefs and projects, and it opens the door to what Grice calls transmitted (but not thereby relative) value. Cybernetic or mechanistic explanations of beings whose essence is to evaluate will not do. The notion of value for such a creature will not be reducible to what is conducive to survival.

Metaphysical transubstantiation: Grice's views on what he called creature construction were first set out in his 1975 American Philosophical Association address.[3] This enormously rich paper cannot be discussed here; the program of creature construction provides a kind of transcendental argument to select those psychological notions which figure in the laws for a given kind of creature. Grice uses it to argue that psychological notions are essentially explanatory, but he can accomodate within this program both privileged access and incorrigibility. His later work takes from this paper the idea that we may plausibly maintain that a capacity and concern for rationality, for justified acceptances, is beneficial for humans. This entitles us to attribute rationality, but only as an accidental property, to humans.

[3] "Method in Philosophical Psychology: From the Banal to the Bizarre," *Proceedings and Addresses of the American Philosophical Association*, LXVIII (November 1975):23–53.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

It does not insure that humans, who are thorough-going in their demand for justification, will succeed. Something more is called for —the operation flauntingly named metaphysical transubstantiation. It is a procedure for the redistribution, but not the invention, of properties. What was accidental becomes essential. Rationality, which attaches only accidentally to *humans,* becomes an essential property of *persons.*[4]

Although there is no new stuff in the world, the world is meant to be richer as a result of this metaphysical routine. If the title has had its full provocative effect, the reader needs some distance from what is supposed to happen to humans in their passage to personhood. More modest examples are needed to illustrate such a metaphysical construction. They may show how it works, and with what justice we could think of this procedure, in general, as a natural disposition.

Mundane examples: I offer three candidates for new entities that come into being via metaphysical transubstantiation. First, sheepdogs. Suppose that there are dogs that have the herding characteristic, that a dog that attached itself to a shepherd exhibited this feature, and that the shepherd came to exploit and develop what he found in this animal. The dog has as an accidental property a capacity and desire to herd. To conceive of this property as essential would be to think of an animal whose job or function it is to herd sheep, and this of course is what we mean by a sheepdog. But, so far, although we have introduced an expression whose use is evident, we could not, says Grice, claim that a new kind of entity had come into being. We would have only "a new way of speaking." It arises from our tendency to reconceive things in terms of their usefulness to us, to shift the emphasis of the properties that characterize them. The natural disposition creates a new kind of entity only if there is legitimate and legitimating motivation. For Grice, a theoretical purpose would be served if we could show that we can extend to the new entity needed explanations. The following motivations seem relevant: first, we would like to be able to extend to the dog those forms of explanation which involve attitudes toward work previously restricted to humans. We need to explain the sheepdog's behavior, using such notions as recognition of a task, eagerness to get to it, and perception that help is needed. We may explain a particular old sheepdog's longevity as we do a person's, in terms of its desire to get its work done. The new

[4] "While each human is standardly coincident with a particular person (and is indeed, perhaps, identical with that person over a time), logic is insufficient to guarantee that there will not come a time when that human and that person are no longer identical, when *one* of them, perhaps, but not the other, has ceased to exist" (*Reply,* p. 102).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

kind of animal is one which essentially cooperates with a human in a job, which can indeed be said to have a job.

Second, we will want to justify treating our co-worker animals in a way different than we treat wild animals. That is, we may want to extend not only principles of explanation, but also practical principles.

It is not in itself the *use* we make of an animal with herding instincts which brings into being a new kind of entity, nor even the training that develops the animal's natural capacities. Although our use and training may bring into the world a particular example of the kind, the construction of the kind is distinct; it is possible only given the shift in concept coupled with a legitimate demand for an entity of the given type. The shift in conception may be what a purposive human requires in order to engage in those activities. What can then be viewed as a necessary presupposition for engaging in those activities backs the claim that a new kind of entity has been constructed.

I want now to refer to, but not examine, two additional examples: pets and tools. Both exemplify the redistribution of properties typical of metaphysical transubstantiation. The legitimating motivation seems purely practical, however. We would like to extend practical principles, possibly, in the case of pets, via an extension of the notion of legitimate expectations. A pet's expectations could then be understood to derive from the kinds of dependency we previously held ourselves responsible for only with respect to humans.

The first tools may have been coincident with found natural objects that exhibited features useful to humans. The practical need to conserve the most apt of these, to keep other humans from casually disposing of them, and to enjoy the advantages of trading for them and improving them, provides practical motivation for the new kind of entity, for the redistribution of properties, such that a tool is *essentially* what is able to do certain work for us. Tools are essentially valuable to us and so whatever practical principles apply to guarding, conserving, and caring for what is valuable get extended to them.

A critic might argue that there is no mystery (and hence nothing metaphysically exciting) in our changing behavior and practices, in extending practical principles, in these two examples. For all that is at issue is the way *we* conceive of an object, not the object's behavior. This is why purely theoretical motivation (as in the sheepdog example) seems stronger. But it is only our conception and our subsequent behavior that is at issue in the case of those social constructions we accept as part of the world: money is a purely conventional unit for the exchange of goods, as is its later substitute in the form of personal checks.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Persons: Grice asserts that there is a rational demand for absolute value which can be satisfied by finding a being whose essence it is to establish and to apply forms of absolute value, and that such a being can be obtained by metaphysical transubstantiation. Presumably, this supplies the rational justification for the move. But what makes the demand for absolute value rational? Grice makes two suggestions, but does not argue for them. The first refers to Leibnizian standards by which a world that contains absolute value would be richer and so better than one that does not. The second suggestion is that we might wish to maintain that it is of absolute value that everyone should, within limits, pursue what is of value to himself. This seems the more promising of the two: if one can argue that there are some essential judgments of relative value which require justification in terms of absolute value, then the demand may be viewed as legitimate or rational.

One may try to exploit the following line: thinking one's ends worth going for is a necessary presupposition of rational action; to have this sense of worth is to have self-respect. There is also a universal human demand that others accord to one some recognition of this claim, that one's ends are worth going for. To grant absolute value to the pursuit of a person's ends is to satisfy that demand.

The basis for legitimating the construction of a new entity seems to have shifted: earlier what Grice required was a theoretical justification, that one would thereby extend principles of explanation which were previously applicable only to a given set of (relatively) primary entities. I suggested we view the examples of tools and pets as operations of metaphysical transubstantiation whose legitimating motivation was practical. The case of humans may be, too, but it is not clear. We sometimes act out of a desire to be, or to become, a certain sort of individual, and our actions are made intelligible by citing that desire. If what I am trying to become is a new kind of entity, my motivation may be said to be practical. Once or if I have succeeded, however, others, to explain my subsequent behavior, must invoke the new kind of entity. But, and now I am in advance of a suggestion I shall shortly make, suppose one just starts with behavior such that what makes it intelligible (to myself or others) is that I have assumed other individuals and myself to be this new kind of entity. If the formal conditions for the metaphysical routine have been satisfied, it is the behavior of the new kind of entity which needs to be explained. What kind of justification is now offered?

Individuals may fairly commonly shift in their sense of what is central to who they are. But when would such a natural tendency legitimate one's reconstitution as a different kind of entity? Two

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

examples may provide Grice with what he wants. They are both highly speculative and sketchy, although the examples themselves are commonplaces: what I propose first is that, when we cooperate with other humans on the basis of trust, our activity is intelligible only if we conceive of ourself and the others as essentially rational. For insofar as I engage in a cooperative venture with other humans on the basis (which I take to be justified) of agreement (tacit or explicit), I must see myself and them as essentially rational. Although in the act of identifying myself with this kind of group I can only construct one given individual instance, myself, I do this via construction of the type.

The underlying ideas are familiar: insofar as I think of the general truth, "we are all alike, we all stand to gain by doing this together," I need impute nothing more than self-interest to myself and others. I can think our capacity to engage in, and act on the basis of, reason-giving to be essentially limited by our self-interest. But, if I believe myself and others to be the kind of creature concerned with standards of acceptability which are not inherently limited by the criterion of self-preservation, if I think we are concerned to find acceptable rules of conduct which are maximally general and which can be justified to other rational beings, then I think of us as that new kind of species—persons.

Let us suppose, as most people now agree, my expectations that others will keep their part of the agreement often go beyond what could be expected on the basis of their calculation of self-interest. Furthermore, to get on with the business of life in a community, some modicum of trust is essential. We are disposed to trust. Our expectations and our trust in others would be at least on the way to being justified if there were such creatures as persons, individuals concerned to find justifiable practical principles, even though metaphysical construction cannot guarantee that the individuals I work with on a given occasion are persons. We can understand then why there is, on the part of humans, a natural disposition to take themselves, and others, to be persons.

To turn to quite a different sphere: it may be that, insofar as a human being commits herself to a theoretical endeavor, to accepting only those theories and findings which can be rationally certified, she has also taken her rational capacities to be essential to herself. Although this kind of example fits many of Grice's ideas, I think that we are on better ground if the practical demand is not made in one's own name alone. If one wished to take a more social view of the enterprise of knowledge than would no doubt have been congenial to Grice, this example might look more like the first.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

HUMEAN PROJECTION

Hume attributed to the mind a tendency "to spread itself upon objects," to project into the world items that, properly or primitively considered, are only features of our states of mind. What served as diagnosis for Hume, however, becomes, under specified conditions, a legitimizing procedure in Grice. It is important to remember that many instances of projecting our wishes and attitudes onto the world do not carry reason or objectivity with them. We find things and people bleak, hateful, awesome, ace, charming, or likeable; we use the vocabulary of our passions to describe the world, but this does not by itself alter what there is.

Grice thinks of Hume's description of the mind's disposition to project as prototypical of the procedure he envisages. He describes two important applications of Humean projection. Their differences arise mainly from the purposes for which Grice deployed them. In the APA address he wished to legitimize the attribution of concepts to creatures which were unavailable to their predecessors. It is part of the story of the development of more sophisticated psychological creatures. It offers a partial explanation of what it is to be more developed, by describing greater internalization or representational capacity in a creature. Internalized expressions included connectives, quantifiers, temporal modiers, mood indicators, modal operators, and names of psychological states like 'judge' and 'will'. In the Carus lectures and the *Reply,* the procedure was used to deliver legitimated concepts.

At the first stage, some initial concept, like that expressed by the words 'or' or 'not' or 'value', is seen as an extrinsic modifier of a psychological attitude (APA), or as an intuitive but unclarified element of our vocabulary (*Reply*). At the second stage, we reach a specific mental state like "or-thinking" (disjoining), "not-thinking" (rejecting, denying), and "value-thinking" (valuing, approving). These states must be manifested in behavioral responses appropriately connected to the initial concept: for example, or-judging *A*, *B*, must generate behavior connected to that of "judging *A* or judging *B*."

At the third stage, reference to these specific states is replaced by a more general psychological verb like 'think' or 'judge', coupled with an operator which appears within, and is restricted to, the scope of the general verb, as in 'thinking *p* or *q*' or 'thinking it valuable to learn Greek'. Only at the fourth stage is the restriction removed, so that the operator can appear in subordinate clauses. Grice warns that stage three will be apparent and not really distinct from stage two, that we will have only a notational variant of an expression, unless

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

certain conditions are met. But these conditions differ in his two essays.

In the APA address, what is required to reach a real third stage, for more than notational change, is that an embedded operator must occur within the scope of another embeddable operator. Thus, '*x* judges (if *A* or *B*, *C*)'. What theoretical purpose is thereby served? Grice speculates: there is the demand for a relatively high grade of inference, in which the passage from premises to conclusion is encompassed, so to speak, within a single thought.

The specific benefit of the procedure described in the APA address is that at stage three, although we cannot yet claim full objectivity for value (what is thought to be of value lacks syntactico-semantical freedom from being "thought-that"), judgments of value are not translatable back into preferences or imperatives or any other approving-volitional attitudes from which they are being constructed. This is arguably due to the more complex linguistic and mental structure in which we find them. It has often been noted that metaethical substitutes of favorings and imperatives do not allow for the kind of argumentation which our value judgments permit. Nontranslatability does not guarantee that there is no counterpart move in the language of approval for every move in the language of value. It does, however, put the onus on the anti-objectivist to make out such a claim.

In the *Reply,* Grice requires that one also reach stage four. Seemingly, it is because "the permissible complexities [can] be made *intelligible*" (p. 98) only via the provision of something like truth conditions for statements within which the operators appear. Legitimate motivation is demanded as well as truth conditions or their analogues; a point or purpose must be served by invoking a new stage. These requirements would surely rule out mere "ways of speaking" and "notational variants," but I think Grice's account of the construction of value needs all the subtleties for which he has himself provided.

I suggest the following: the requirement for reaching stage three of the APA address rules out mere notational variants and the formal satisfaction of stage four, unaccompanied by motivational justification, gives us "ways of speaking." One may correctly call a movie 'ace', and this expression may be more than a notational variant. But such an expression, or projection, introduces no new kinds of entities.

In the Carus lectures and the *Reply,* truth conditions or some analogous valuation is called for to reach stage four. The likely candidate for accreditation is the philosophical story Grice himself tells.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

This would permit us to attribute to persons, at least, value as members of a kind, where the statement that persons are of value will be detachable from its psychological attitude.

Of course, one will think that many other things are of value. To get to the objectivity of those judgments, Grice invokes the following principle: Humean projection, if performed by a good person in conditions of freedom, allows for the transmission of (absolute, nonrelative) value from the person to what seems to her valuable. Grice thought his principle an analogue of the Aristotelian principle, "what seems good to the good man is good." Its rationale seems to depend on the following elements of Grice's story: as rational, we demand justification of our projects and attitudes; the standards for acceptance do not, however, come ready-made, but must themselves be constructed. Persons who work at their job, who are concerned to find principles acceptable to rational beings, get the benefits of their job: what seems good to them is.

I want to raise for discussion several difficulties: let us suppose that we are convinced by Grice's philosophical examination of value and finality; we are thus led to think that value of a kind can be attributed to persons, and that some humans do reconstitute themselves as persons. *Can* we legitimately think them of value? Value was a suspect term, one which had to be put through the mangle of Humean projection in order to be legitimated. Now, Humean projection in the APA address would allow me at least to think persons to be of value (and my judgments would, of course, motivate me) but the statement would have no claim to objectivity. It seems that only my own metaphysical transubstantiation will permit me to reap the benefits of objective attribution of value. That by itself may not be undesirable, but, in order to play the game in its second application, in order to reap the benefits, I must myself be of absolute value. Now, I am not sure that this is viciously circular, but I should like to be convinced that it is not.

I shall raise two other very general questions at the APA meeting this December: Does construction have to be social? Does Grice adequately answer the skeptic who would happily endorse the constructions but withhold the claim to objectivity?

JUDITH BAKER

York University

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ON GRICE ON LANGUAGE*

H. P. Grice's philosophy of language shares with Oxford ordinary-language philosophy an emphasis on language as a social phenomenon. We say things to others to achieve ends that are possible only through language use and only through cooperation with others. But, unlike most Oxford philosophers of language, Grice was not satisfied with classifying bits of language and collecting the odd specimen. He strove for a theory, a theory that would do justice to the phenomenal richness of human languages while unraveling the philosophical perplexities about meaning, truth, and denotation.

I. INTRODUCTION

His work in this area divides into three parts: the writings on meaning have been published at intervals over a 30-year period, and have spawned variants and objections and commentaries. His work on conversational implicature was publicly presented in the William James Lectures at Harvard in 1967, but was completely unpublished until 1975[1] when a somewhat misleading piece was revealed, the theory in its entirety not appearing until just after his death in *Studies in the Way of Words*. These ideas, too, were widely-known and discussed, but, in the absence of official texts, the discussions have been less focused. In contrast, his work on truth and denotation has been little discussed, partly because it has not been published, though it was presented in a series of lectures at the University of Illinois and at a national summer institute at the University of California/Irvine in 1971.

The best-known application of the idea of conversational implicature is Grice's exploration of the possibility of defending the classical logician's claim that the material conditional has the same meaning as the nonsubjunctive uses of 'if-then' in English. But this was not the first use Grice made of the idea, nor was it the one he regarded as most important. The idea first appears in print in his 1961, "The Causal Theory of Perception."[2]

* To be presented in an APA symposium on the Thought of Grice, December 28. Judith Baker will be co-symposiast, and Robert Stalnaker and Richard Warner will comment; see this JOURNAL, this issue, 505–513, 526/7, 528/9, respectively, for their comments.

[1] "Logic and Conversation," in *The Logic of Grammar*, Donald Davidson and Gilbert Harman, eds. (Encino, CA: Dickenson, 1975), pp. 64–75. Page references are to the reprinting in *Studies in the Way of Words* (Cambridge: Harvard, 1988).

[2] Reprinted in *Studies in the Way of Words*.

0022-362X/89/8610/514–525 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Grice begins there by noting that causal theories of perception had received little attention (a situation his article helped remedy) and his goal was to formulate a causal theory of perception (CTP) that, in his characteristically guarded phrase, "is, if not true, at least not too obviously false" (224). The version of CTP he chose to defend is closely related to H. H. Price's and includes the "thesis that 'I am perceiving M' is to be regarded as equivalent to 'I am having (or sensing) a sense-datum which is caused by M' " (225). What concerns Grice is not the causation condition but the problems involved in defending sense data.

He regards as unsatisfactory attempts to prove the existence of a special class of objects via arguments from illusion, as well as Moorean approaches which give instructions on how to identify the sense data in one's vicinity. He believes a more promising strategy is not to identify sense data as objects, but to provide paraphrases for sentences containing apparent references to sense data into sentences that contain no such references. The particular general strategy he envisions is to translate sentences such as 'I have a red sense datum' via sentences such as 'It looks to me as if something red is present', where the something can be a physical object since the 'looks' locution provides a hedge.

He remarks that the program of carrying out such a reduction is by no means easy and that various expressions that have been proposed for the key role ('looks/feels/sounds', 'seems', 'appears') have "more or less subtle differences between them which would have to be investigated" (226). What concerns him is not the details of the reduction but a general objection, namely, that "there would be something at least *prima facie* odd about my saying 'That looks red to me' (not as a joke) when I am confronted by a British pillar-box in normal daylight at a range of a few feet" (227). The objector concludes that the oddity is due to the falsity of the statement; Grice counters by locating the oddity in some other linguistic factor.

In the James lectures, Grice lists similar arguments that bedevil various philosophers' claims and analyses. For example, Norman Malcolm argued that Moore was wrong to claim that he knew there was a hand before him, because it would be very peculiar to hold up a hand and exclaim: "I know this is a hand." Claims about remembering also are problematic in the same way. It is peculiar, under normal circumstances, to say that you remember your name, but is it false that you remember it? It can also be argued in a similar vein that, if I pick up a coffee cup under normal circumstances, it would be odd to say that I tried to pick it up, or that I voluntarily picked it up, or that I carefully picked it up. And in these cases the conclusion contem-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

plated is that therefore it is false that I tried to pick it up, that I voluntarily picked it up, or that I carefully picked it up.

These intuitions of oddness seem solid, but is the further inference to falsity justified? One way to buttress the inference is to argue that proper use of the sentences in question includes *not* saying them under the circumstances envisaged. That is, one does not know how to use the sentence, 'He tried to lift the coffee cup', if it is said when the person had no difficulty and no reason to think there would be difficulty in lifting the cup. If one accepts the theory that use is meaning, then it is part of the meaning of 'He tried to lift the coffee cup' that there is difficulty involved, or there is believed by someone to be difficulty involved, in lifting the cup.

II. THEORY

Grice's theory of conversational implicature rejects these arguments and the associated theory of use as meaning.

> . . . indeed, the precept that one should be careful not to confuse meaning and use is perhaps on the way towards being as handy a philosophical vademecum as once was the precept that one should be careful to identify them.[4]

On his theory, the conventional meaning of a sentence, and what is said on a particular occasion of utterance, are only part of the total linguistic force. As mentioned earlier, it is central for Grice that language use is a form of rational action, and, in choosing a particular sentence to utter, the speaker is choosing that sentence *rather than a number of alternatives.* In our example, if the speaker says 'He tried to lift the coffee cup', then he has chosen that sentence rather than 'He lifted the coffee cup'—if he indeed succeeded in lifting the cup, then it would take some special reason to choose the longer sentence.

Thus far we are opposing one intuition—that it is false to say that he tried to lift the cup—with another—that it is odd but true. What Grice offered in the James lectures was a theory to classify the kinds of oddness and to back the second kind of intuition.

He begins by observing that conversations are not random, unrelated remarks, but serve purposes. If they are to continue, then they must involve some degree of cooperation and some convergence of purposes. (Note that convergence does not mean identity, and purposes may change during the conversation.) Otherwise, at least one of the parties would have no reason to continue the conversation, and we are presuming that the participants are rational agents.

> We might then formulate a rough general principle which participants will be expected (ceteris paribus) to observe, namely: Make your conver-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

> sational contribution such as is required at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged. One might label this the Cooperative Principle (26).

Under the general umbrella of the "cooperative principle," Grice distinguishes more specific maxims and submaxims. He notes that there are overlaps among the principles—being relevant may well entail not providing any more information than is required.

And the principles may often conflict. If one has partial evidence for an answer that is as informative as is required, one is caught between the maxim enjoining one to be informative and one cautioning against saying things for which you have insufficent evidence.

It is important that, although the specific formulations given here apply in many cases only to linguistic exchanges, Grice views these principles—suitable generalizations—as applying in any cooperative venture, linguistic or not. He gives some illustrations of this idea, though he does not attempt to formulate the more general principles.

> Relation. I expect a partner's contribution to be appropriate to the immediate needs at each stage of the transaction. If I am mixing ingredients for a cake, I do not expect to be handed a good book . . . (28).

He remarks further that he would "like to be able to show that observance of the Cooperative Principle and maxims is reasonable (rational) . . . (29–30)," because participation in talk exchanges will be fruitful only if the exchange conforms to the maxims. He notes that, however,

> Whether any such conclusion can be reached, I am uncertain; in any case, I am fairly sure that I cannot reach it until I am a good deal clearer about the nature of relevance . . . (30).

Each step in a conversational exchange can be analyzed in terms of whether it conforms to the maxims or not. There are four possibilities: in the most straightforward case, all maxims are obeyed; in the most devious case, a maxim is disobeyed but without the knowledge of the other participant. A different case is one where a participant overtly opts out, perhaps by saying, "No comment." I have already mentioned the possibility of a conflict of maxims, in which case there is a cooperative reason for not following all of them. Yet another case is where a maxim is flouted—it is disobeyed not secretly nor by avowal, but by a clearly nonconforming performance.

What interested Grice, however, were the cases in which a maxim appears to be violated but where, upon further reflection, the audience can figure out, on the assumption that the speaker is being generally cooperative, how the speaker is being subtly cooperative.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

> How can his saying what he did say be reconciled with the supposition that he is observing the overall Cooperative Principle? This situation is one that characteristically gives rise to a conversational implicature; and when a conversational implicature is generated in this way, I shall say that a maxim is being *exploited* (30).

Maxim exploitation can be minor, obvious, and transparent, or it can be major and surprising. One example Grice gives is roughly:

A. I am almost out of gas.
B. There is a station around the corner.

In this case, *B* has not *said* that the station is a gas station, rather than a television station, and has not *said* that the station is open. But *B* has conversationally implicated both of these, for otherwise the response is not relevant to the conversation. But, in this case, the step to the implicatures are so small and immediate and natural that we would not normally even notice that they occur.

To put these matters in a larger perspective, Grice distinguishes three elements within the total signification of an utterance:

a. what is said
b. what is conventionally implicated
c. what is conversationally implicated

We saw with regard to such expressions as 'try', 'seem', and 'remember' that part of the point of distinguishing conversational implicatures from meaning is that certain implications of an utterance were to be analyzed as depending on factors other than meaning. But, if the notion of conversational implicature is to be something other than an ad hoc device, we need some methods for telling when an implicature is a conversational one and when it is conventional. Otherwise, there is no principled way of arguing that, if someone says Jones seems to see a purple petunia, this does not *mean* that he does not see one, but only conversationally implicates that he does not.

III. DEFINITION

First, we need a definition of conversational implication with which to work.

> A man who, by (in, when) saying (or making as if to say) that p has implicated that q, provided that
> (1) he is to be presumed to be observing the conversational maxims, or at least the cooperative principle;
> (2) the supposition that he is aware that, or thinks that, q is required in order to make his saying or making as if to say p (or doing so in *those* terms consistent with this presumption; and

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(3) the speaker thinks (and would expect the hearer to think that the speaker thinks) that it is within the competence of the hearer to work out, or grasp intuitively, that the supposition mentioned in (2) is required (30/1).

Note that the characterization is intended to mark off among the class of implicatures those which are conversational as opposed to conventional. Although the definition only requires that the speaker think the hearer can intuitively grasp the required supposition, Grice imposes the stronger requirement a few lines later:

> The presence of a conversational implicature must be capable of being worked out; for even if it can in fact be intuitively grasped, unless the intuition is replaceable by an argument, the implicature (if present at all) will not count as a conversational implicature; it will be a conventional implicature (31).

There may seem to be a conflict between the two statements, but I take it that the second is an *in principle* requirement. In other words, it suffices if the hearer can intuitively grasp the implicature so long as we conversational theoreticians can supply the rigorous argument. I shall discuss later some difficulties with the requirement; for the moment, I shall simply make explicit the analysis of an example given above.

If under normal circumstances a speaker utters the sentence, 'Paul tried to lift the coffee cup', and we assume that the speaker is following the cooperative principle, the hearers must ask themselves why the speaker did not say, 'Paul lifted the coffee cup'? The latter is a shorter and simpler sentence, and, normally, if one is in a position to know that someone tried to lift a cup, one is also in a position to know whether he succeeded. Moreover, the audience can be expected by the speaker to know of the existence of the shorter sentence and of the epistemic regularity cited. Thus, if Paul indeed lifted the cup, the speaker would have known that fact and would, under the cooperative assumption, have informed us, and so we can infer by modus tollens that the speaker does not believe the cup was lifted, and, furthermore, since he was in a position to know, we can infer that he believes the cup was not lifted.

IV. TESTS AND CONSEQUENCES

Grice does not give a test for conversational implicature, other than the characterization cited; but he does develop a list of features the conversational implicature must possess. This list has been taken by numerous writers—some critical of Grice, some applying the concept—as constituting a set of *definitive* features. This is due to the fact that his ideas on conversational implicature were known mainly

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

from "Logic and Conversation," which was constituted by Lecture II of the James Lectures, while a significant disclaimer appeared early in Lecture III, which was printed in 1978[3] but seems to have been less widely read.

> . . . I was not going so far as to suggest that it is possible, in terms of some or all of these features, to devise a decisive test to settle the question whether a conversational implicature is present or not . . . Indeed I very much doubt whether the features mentioned can be made to provide any such knock-down test, though I am sure that at least some of them are useful as providing a more or less strong prima facie case in favor of the presence of a conversational implicature (42/3).

The first, and most diagnostically useful, feature is cancelability. Since the demonstration of the implicature depends on several premises, by denying one of those premises, either overtly or indirectly, the path to the implicature can be blocked. In our example, if the speaker adds a second sentence thus:

> Paul tried to lift the coffee cup. I'm not going to reveal whether he succeeded.

then the implication is cancelled, because we know the speaker is not being fully cooperative. In other cases, a speaker may indicate that he has no further information:

> Paul tried to catch the 3:05 train, but I doubt that he made it.

suggests that he did not catch the train, but does not *implicate* that he did not.

The list of features is not as much help in distinguishing genuine conversational implications from conventional implications as one might think at first glance. One presupposes that distinction, while two others depend heavily on the technical, theoretical notion of *what is said.* The last is not a test, but a useful reminder that indeterminacy or vagueness need not rule out a candidate for conversational implicature.

V. OBJECTIONS

The theory of conversational implicature is undoubtedly Grice's mostly widely-known and accepted contribution to the philosophical enterprise, but it is not without problems. Criticisms point in several directions. Some object to the lack of detail and specificity—the theory is seriously incomplete, they note. The other objections question some of the details that are given—usually they raise questions about some of the maxims.

[3] "Further Notes on Logic and Conversation," in P. Cole, ed., *Syntax and Semantics: Pragmatics, vol. 9* (New York: Academic, 1978), pp. 113–127.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

I shall consider three of the objections:

Objection 1. There is often no common goal in a conversation, and thus such conversations do not satisfy the cooperative principle and thus, *on Grice's account,* generate no conversational implicatures.

Objection 2. The account allows too many propositions as implicatures.

Objection 3. On some plausible accounts of reasoning, the maxims do not enter.

VI. REPLIES

Objection 1: No common goal. Asa Kasher[4] raises this objection to persuade readers that a different set of maxims should be substituted for Grice's. Conversations certainly differ on the degree to which there is a shared purpose. The most convergent are those where the participants have a common specific goal that is paramount, e.g., on both getting to the airport as quickly as possible. In other cases, there may be a nonspecific goal that is paramount with divergent specific ways they wish to realize the goal, e.g., George and Mike may have as their predominant goal that they have dinner together at a restaurant; but George may have a very strong preference for a Mexican restaurant and a strong aversion to Greek food, while Mike has the opposite desires.

It is important to distinguish these cases, which seem to fall naturally within the general framework of the cooperative principle, from others where there is a more elusive shared goal. Imagine that *C* wants to buy a car for less than $9,000, and that *D* has a car which would satisfy *C*'s other requirements and which *D* wants to sell for more than $10,000, and both know this. We can conjure up a shared goal: "*C* buys *D*'s car at a mutually agreeable price." It may be stretching ordinary usage to describe them as cooperating, though they both want the conversation to continue. *C* would not talk to *D* if he did not think he could achieve his goal, and similarly for *D*. Thus, the purpose of the conversation is to persuade the other party to change his goal. The resultant conversation may be replete with conversational implicatures, and it may be quite rational for each party to continue the conversation, for a while at least.

What is required for rational agents to continue a conversation is not that they share a fully specified goal, nor even that their goals interact as George and Mike's, but that their goals interact with their situations so that each has an expectation of benefitting from the conversation. Some conversations are similar to poker games in which it may be rational for each of us to play with the expectation of winning, even knowing that we cannot both win.

[4] "Conversational Maxims and Rationality," in *Language in Focus,* A. Kasher, ed. (Dordrecht: Reidel, 1976), pp. 197–216.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Objection 2: The account allows too many implicatures. Grice anticipates this objection and requires that the implicatures be nontrivial. For example, suppose Karl seems to utter sincerely the statement that Paul tried to lift the coffee cup. His saying it would make no conversational sense if he did not believe it, but that he believes it is not a logical consequence of what he said. Since the hearer can be expected to draw the conclusion, it meets the requirement for a conversational implicature. Grice gives no measure of triviality, but clearly envisions at least excluding schematically predictable candidates for conversational implicature. In the specific instance, whenever someone says p, there is a potential conversational implicature that they believe p. Thus, this potential implicature is trivial and thereby excluded.

A second kind of case to be ruled out concerns mutual knowledge or general information.[5] An example adapted from Grice is:

A: Has Smith been dating anyone?
B: He has been visiting New York every weekend.

> The implicature, according to Grice, is that Smith has been dating someone in New York. . . . The problem is that there are other things which the theory seems to imply that B implicates. For example, it seems to imply that B implicates that New York is a place where people live (*ibid.*, p. 16).

This particular implicature could be ruled out by excluding anything that was already common knowledge between the participants.

Unfortunately, that exclusion does not suffice to keep out all unwanted candidates for implicature, because new knowledge may arise from the combination of background knowledge and the intended implicature. Imagine that the previous conversation takes place in Jamaica where *A*, *B*, and Smith all live. Since *A* and *B* both know that the only way to go to New York on the weekend is via Air Jamaica, *B*'s response is consistent with his obeying the maxims, only if he believes that Smith flies Air Jamaica every weekend, and he can expect *A* to come to that realization. Moreover, it was not common knowledge before *B*'s remark that Smith flies Air Jamaica every weekend. Thus, that proposition meets the conditions for being a conversational implicature (*ibid.*, p. 17).

One way of responding to the criticisms would be to strengthen the third condition for conversational implicature and require not merely that the speaker *think* that the hearer will work out the supposition in the second clause, but that the speaker *want* the

[5] Cf. Andrew S. McCafferty, *Reasoning about Implicature: A Plan Based Approach* (doctoral dissertation, University of Pittsburgh, 1987).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

hearer to work out the supposition. In our last case, *B* may think, even expect, that *A* will work through to the conclusion that Smith flies Air Jamaica, but may be quite indifferent as to whether *A* does so. But *B* presumably does want *A* to come to the conclusion that Smith is dating someone.

Objection 3: On some plausible accounts of reasoning, the maxims do not enter. This objection does not question the phenomena of conversational implicature, only Grice's account of it as based on his maxims. One of Grice's examples is a thesis advisor whose entire recommendation for an advisee is that he has good handwriting and is punctual. The reasoning of the hearer might be filled out as follows, according to a Gricean account:

1. *A* is a respected member of the philosophical profession.
2. *A* wishes to continue to enjoy the regard of the philosophical profession.
3. If *A* were not to cooperate in evaluating a candidate about whom *A* had information, *A* would be violating the expectations of the profession.

Therefore,

4. *A* has good reason to cooperate in this inquiry.
5. *A* has no overriding reason not to cooperate.

Therefore,

6. *A* is cooperating in this conversation.

Therefore,

7. *A*'s remark is intended to be informative.

The objection, also due to McCafferty, is that one can alternatively, and more parsimoniously provide an alternative course of reasoning which simply omits (5) and (6)!

In general, the objection goes, where one has reasons to believe that someone is cooperating, those reasons suffice to generate an explanation of the implicature without detouring through the higher level cooperative principle.

Notice that, as a matter of practical reasoning, it is important that we have reasons to think that our fellow conversationalist is being cooperative in some form to some degree. If we ask a stranger for directions, it is relevant to believe that he would rather admit not knowing how to get to our destination than invent an answer to hide his ignorance.

Grice never discussed the reasoning that might lead one to the conclusion that someone is being cooperative, and so we have no direct knowledge of how he might reply to this line of objection. His

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

attitude was probably linked to his general ethical views, however. The kinds of reasoning suggested as sparser arguments bypassing the cooperative principle are utilitarian in character, and he was no friend of utilitarianism. I suggest, therefore, that he would have argued that the cooperative principle ought to be a governing principle for rational agents on Kantian grounds. Thus, for rational agents of the kind he envisioned, the account in the James lectures would be correct, for they would follow the cooperative principle on moral, *not* on practical or utilitarian grounds.

VII. IMPLICATURE, MEANING, AND REDUCTION

We have been exploring part of Grice's writings on the philosophy of language, but I want to consider another part in closing. The discussion of conversational implicature contrasts pragmatic contextual matters with the fixed aspects of semantics. I think that this contrast is often overstated in discussions of Grice's philosophy of language. Although distinguishing pragmatic matters of implicature from conventional matters of meaning is central to his discussions of several topics, the difference was not, in his mind, so clear or clear-cut as in the minds of commentators and followers. I think that this is connected with the natural, but I believe erroneous, construal of his views on meaning as being essentially reductionistic.

His characterization of meaning proceeds from the very pragmatic and context-bound concept of what a speaker means by a particular utterance through many stages to the concept of timeless sentence meaning in a language. There is, however, a gap between the claim that intentions and beliefs determine the semantics of the language spoken by a group and the claim that statements of semantic fact can be individually reduced to statements about intentions and beliefs. A typical Gricean analysis of applied timeless sentence meaning relies on the instances of utterer's meaning within the relevant group.

Utterer's meaning in turn is cast in terms of intentions to produce various propositional attitudes among the hearers of the utterance. Such analyses are susceptible to numerous infamous counterexamples, the details depending on the specific propositional attitudes chosen, which I shall not repeat.

Even if the program were successful to this point, the problem would remain of how to deal with the propositional attitudes. Grice was uncertain on this point, though at times he explored the possibility of reducing propositional attitudes to less sophisticated and less cognitive attitudes, with the eventual end in view of gradually reducing those to yet more primitive and extensional idioms. There are problems at every proposed level, many of them lurking unarticulated because the specific reductions have not always been suggested.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Still, it is a major further step to turn the insight that intentions somehow determine language into atomistic reductions of meaning, sentence by sentence. An alternative view is that the relation between intentional and intensional concepts might prove to be one of mutual interrelationship and interdependence. The Gricean analyses, even if one could find versions that were ultimately immune to counterexamples, need not be seen as reductions, but may be viewed rather as axiomatic relations among concepts. Grice himself seems to have viewed them thus at times.

He also, at least at a later stage, tended to view the relations as theoretical ideals, similar to point masses in mechanics. Consider, for example, his preferred method of handling the question of how many iterations of propositional attitudes are to be embedded in the definitions of various kinds of meaning. How many iterations should we require of the form, 'the speaker intends the hearer to recognize that the speaker intends . . .'? Grice opted for the answer that we should in our theory require an infinite array of such iterations, but that we should recognize in applications that we could at most expect the absence of contrary intentions. Again this is an example of how his approach was, or at least became, one of providing explanations rather than procedures for elimination.

Moreover, his general metaphysical approach, unfortunately never articulated to any great extent in print,[6] was at odds with some main currents of contemporary thought. Given an analysis of a concept in terms of others, he would regard this as a legitimation of the concept, a justification for its use, rather than evidence that it was eliminable and suspect! His goal was not to show that meaning and other semantic concepts were unproblematic because they were eliminable, but to show that they were unproblematic because they could be explicated. His preference in metaphysical landscapes ran more to luxuriant gardens than to desert scenes.

RICHARD E. GRANDY

Rice University

[6] Some of his views emerge in "Reply to Richards," in *Philosophical Grounds of Rationality: Intentions, Categories, Ends*, Richard E. Grandy and Richard Warner, eds. (New York: Oxford, 1986), pp. 45–108.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ON GRANDY ON GRICE*

MY COMMENTS will focus on H. P. Grice's work in the philosophy of language, and so on Richard Grandy's contribution to this symposium. Grandy's paper concentrates on Grice's notion of conversational implicature and on the theory of which that notion is the centerpiece. After giving a brief exposition of Grice's theory of conversation, Grandy considers three objections that have appeared in the literature, and discusses how Grice might have replied to them. He concludes with some intriguing but brief remarks: first, about the contrast between pragmatics and semantics; and, second, about Grice's general views about philosophical analysis.

I find little to quarrel with in Grandy's exposition of Grice's motivation, his theory of conversation, and the strategy of argument in his applications of the theory to philosophical problems. I also agree with what I take to be Grandy's implication that the objections he discusses are not serious threats to the Gricean program. But the objections do direct our attention to some interesting questions that I want to discuss: first, concerning the cooperative principle and the presumption that speakers can be expected to observe it; and, second, concerning the relation between conversational implicature and speaker meaning.

Grandy speculates that Grice might have linked the presumption that speakers observe the cooperative principle to his ethical views, suggesting that "rational agents of the kind he envisioned . . . would follow the cooperative principle on moral . . . grounds." Whatever the merits of this as a speculation about what Grice might have said, I shall argue that an appeal to ethical theory would be unnecessary and a mistake. The cooperative principle, and the maxims it justifies, play a role in the conversations of the most unsavory conversationalists, of those who have little confidence in each other's moral character. I think one can and should explain the role of cooperation in conversation independently of the assumption that rational persons are motivated by any kind of moral principle.

Grandy distinguishes Grice's discussions of the analysis of meaning from his account of conversation implicature, but I think it is important to see these as a part of a single project. As I understand the

* Abstract of a paper to be read in an APA symposium on the Thought of Grice, December 28, commenting on papers by Judith Baker and Richard Grandy; Richard Warner will also comment. See this JOURNAL, this issue, 505-513, 514-525, and 528/9, respectively.

0022-362X/89/8610/526/7
© 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

intended notion, conversational implicature is a kind of speaker meaning, a kind distinguished by the source of the expectations in terms of which speaker meaning is defined. The definition of conversational implicature should reflect this relation, and, if it did, then one of the objections Grandy discusses would not arise.

I agree with Grandy that the contrast between contextual factors and literal semantics "is often overstated in discussions of Grice's philosophy of language." I think the Gricean framework allows for a more fluid and complex relation between semantics and pragmatics, and I shall suggest some ways in which the conversational maxims might be used in explanations of what is literally said, as well as of what is merely implicated.

On Grandy's concluding remarks about reduction and explication, I shall press for more details. The analyses that Grice sought and gave certainly seemed to be definitions in the strictest sense—definitions of the same form that is sought and given by proponents of philosophical reduction. But, Grandy says, "his goal was not to show that meaning and other semantic concepts were unproblematic because they were eliminable, but to show that they were unproblematic because they could be explicated." But is this more than a difference of attitude toward one's philosophical explanations? Would the strictest reductionist not agree that an analysis of a concept legitimates it, provides a justification for its use? How are the philosophical explications favored by Grice different from the philosophical analyses of those who see them as eliminative reductions?

ROBERT STALNAKER

Massachusetts Institute of Technology

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

REPLY TO BAKER AND GRANDY*

I SHALL RESPOND primarily to Judith Baker's insightful discussion of H. P. Grice's work in metaphysics and ethics. I shall begin, however, with some brief general remarks about the nature of Grice's later work and shall also comment briefly on meaning and implicature.

A good deal of the later work is concerned with the concept of rationality. This is no surprise when one realizes the centrality of that concept to Grice's thought. To begin with, the concept underlies Grice's approach to metaphysics and ethics. Metaphysics for Grice *just is* the account of the most general and noncontingent constraints on rational thought. Metaphysics so conceived yields an understanding and legitimation of the concept of final causes as well as an explication, via metaphysical construction, of the concept of a person. These two concepts are essential elements in the explication and defense, sketched in the *Reply to Richards,* of the notion of absolute value. Turning from metaphysics and ethics to philosophy of language, the concept of rationality provides the foundation for the work on meaning and the philosophical psychology within which that work proceeds. Meaning and associated phenomena paradigmatically are, or arise out of, cooperative efforts between rational beings with a common interest in communication.

The distinction between what is meant and what is implicated is intended as part of a philosophically illuminating description and explanation of rational, cooperative efforts at communication. The distinction keeps us from mistaking what is typically implicated for what really is part of what we mean. Grice, as is well known, made good use of this distinction—in, for example, his defense of the causal theory of perception. The distinction between meaning and implication has sometimes been criticized for presupposing the analytic/synthetic distinction: if there is a distinction between meaning and implicature, then some connections hold as a matter of meaning, and some do not. Grice did accept the analytic/synthetic distinction and, in fact, thought it essential. An acceptable account of the distinction was supposed to emerge from the general account of rationality—from metaphysics. Grice's idea was that, at the ideal limit of full, explicit development, a theory defined, or contained

* Abstract of a paper to be read in an APA symposium on the Thought of Grice, December 28, commenting on papers by Judith Baker and Richard Grandy; Robert Stalnaker will also comment. See this JOURNAL, this issue, 505-513, 514-525, and 526/7, respectively.

0022-362X/89/8610/528/9
© 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

resources to define explicitly, all but the primitive concepts it employed. These definitions would express analytic truths. Grice thought it a requirement on rational thought that theorizing be guided by this, perhaps forever unobtainable, ideal. I say this is Grice's idea, and it is. He returned to it many times in conversations about the analytic/synthetic distinction. But he was never comfortable with the idea.

Let us turn to metaphysics and ethics. Baker asks if metaphysical construction yields objectivity. She notes that Humean projection allows us to think that persons are of value, but, she adds, "the statement would have no claim to objectivity." She continues: "It seems that only my own metaphysical transubstantiation will permit me to reap the benefits of objective attribution of value." And this, she worries, may be viciously circular, for "to reap the benefits, I must myself be of absolute value."

I think that neither Humean projection nor metaphysical transubstantiation, singly or together, yield objectivity. Rather, objectivity arises from the fact that, as Grice puts it in the *Reply*, "there is a rational demand for absolute value." As Baker points out, this demand most plausibly arises from the "need to appeal to absolute value in order to justify some of our beliefs and attributes with regard to relativized value." It is a condition of being rational that we think some things of absolute value. Humean projection and metaphysical transubstantiation are part of a transcendental—or, as Grice preferred, metaphysical—argument that shows us how it is possible for there to be a being that *must,* insofar as it is rational, think that it is of absolute value. We are to recognize this being as ourselves—*qua* persons. This shows us that rationality demands the acceptance, but not the truth, of our being of absolute value. But then, as Grice remarks, this feature is not "a defect if one can go on to say that *this* kind of demand for acceptance is sufficient to confer truth on what is to be accepted." Grice more or less took this point for granted: that is, he assumed that, if one must, insofar as one is rational, think something of value, then it is *true* that it is of value. He did not, as far as I know, ever offer a reason why this was so.

But it is so. Or at least so I claim—more boldly and baldly than Grice would ever have cared to. But then my role with Grice was always of advance scout—staking out positions and reporting to the general. I see no reason to play any other role here.

RICHARD WARNER

University of Southern California

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

APPLYING KRIPKE'S THEORY OF TRUTH*

SAUL Kripke's "Outline of a Theory of Truth"[1] is an extraordinary document. Nearly everyone can agree that it is a profoundly important piece of work, but scarcely anyone can agree where its importance lies. Its subject is the liar paradox, which is a problem of the utmost philosophical urgency. Unless we can overcome the liar paradox, we cannot develop a consistent theory of truth; and without a consistent theory of truth, we cannot obtain a satisfactory understanding of the relation between our thought and language and the world around us. Here I would like to offer both a negative thesis—that something that looks like an important feature of Kripke's theory is really unimportant—and a positive thesis about what is deeply significant in Kripke's work.

The negative thesis is that truth-value gaps are a red herring. They figure prominently in Kripke's text, but they are not essential either to understanding the mathematical construction Kripke develops or to obtaining a correct understanding of truth.

The positive thesis, which arises out of Kripke's careful investigation of the everyday, nonphilosophical, nonparadoxical use of the notion of truth, is that what we require of a satisfactory theory of truth is not that it give us Tarski's schema

(T) $\ulcorner\phi\urcorner$ is true (in English) if and only if ϕ.

but that it yield these four rules of inference:

(R1) From $\ulcorner\phi\urcorner$ one may infer $\ulcorner\ulcorner\phi\urcorner$ is true$\urcorner$.
(R2) From $\ulcorner\ulcorner\phi\urcorner$ is true$\urcorner$ one may infer $\ulcorner\phi\urcorner$.
(R3) From $\ulcorner\neg\phi\urcorner$ one may infer $\ulcorner\ulcorner\phi\urcorner$ is not true$\urcorner$.
(R4) From $\ulcorner\ulcorner\phi\urcorner$ is not true$\urcorner$ one may infer $\ulcorner\neg\phi\urcorner$.

Using Kripke's methods, it is possible to develop a theory of truth which validates (R1)–(R4), and, indeed, it is possible to develop such a theory for a language within the language itself, without recourse to an essentially richer metalanguage.

To learn how to develop the semantics of a language without requiring a richer metalanguage would be an accomplishment of the

* To be presented in an APA symposium on Kripke's Theory of Truth, December 29. Stephen Yablo will comment; see this JOURNAL, this issue, 539–541. Thanks to Marian David and Scott Sturgeon for helpful comments.

[1] This JOURNAL, LXXII, 19 (November 6, 1975): 690–716.

0022-362X/89/8610/530–539
© 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

greatest importance. It would overturn the presently accepted doctrine, propounded by Tarski:[2]

> *Tarski's Commandment.* Thou shalt not develop the semantics of a language within the language itself, but in an essentially richer metalanguage.

The restrictions this commandment imposes are severe, for it implies that, having no metalanguage richer than English, we cannot develop a semantic theory for English, or for any natural language. We can develop a semantics for a fragment of English which includes the terms we use when we talk about chemistry, but we cannot extend the account to include the terms we use to talk about semantics. The terms we use when we talk about language lie outside the realm of linguistic inquiry, and, for similar reasons, the concepts we use when we think about thought lie outside the realm of psychological inquiry.

Now, Kripke accepts Tarski's commandment, but we can, I shall claim, apply Kripke's methods in a manner different from what Kripke himself had in mind, in order to break free of the restrictions of Tarski's commandment. Perhaps someday we can use these methods to obtain a unified scientific understanding that comprehends nature, human language, and human thought as parts of a connected whole.

The mathematical core of Kripke's paper is the production of a model in which there are truth-value gaps and in which a sentence $\ulcorner\phi\urcorner$ is true, false, or truthvalueless accordingly as $\ulcorner\ulcorner\phi\urcorner$ is true$\urcorner$ is true, false, or truthvalueless; such a model is called a "fixed point." Here truth-value gaps are handled by Kleene's strong 3-valued logic,[3] so that, for example, a disjunction is true if either disjunct is true, false if both disjuncts are false, and truthvalueless otherwise.

Kripke is remarkably reticent about the philosophical import of his formal construction. I suspect that this is both willful and wise. If he had specified a philosophical purpose, people would have looked only at how well the construction fulfilled that purpose, and not at other purposes it might serve.

His most explicit statement of purpose (*ibid.*, p. 80n) is that he intends to represent the intuitive response of a nonphilosophical speaker who has seen the *simple liar* ('This statement is false'.) but

[2] "Der Wahrheitsbegriff in den formalisierten Sprachen," *Studia Philosophica* 1 (1935): 261–405 [in Tarski's *Logic, Semantics, Metamathematics,* Joseph Woodger, trans. (Indianapolis: Hackett, 1983), pp. 152–278].

[3] *Introduction to Metamathematics* (New York: Van Nostrand, 1952), pp. 332–340.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

who has not seen and reflected upon the *strengthened liar* ('This statement is not true'.). Since most people are shown the two forms of the paradox one right after the other, so that the intuitions Kripke is trying to codify persist for only a few moments, this is an excessively modest aim. But even this modest aim is not well-fulfilled, since there is no stage at which the reasoning of ordinary speakers exhibits the pathologies characteristic of 3-valued logic.

I would like to look at how well Kripke's construction can be used for another purpose, quite different from anything Kripke himself proposed. The logical paradoxes show that our naive understanding of truth, which includes the acceptance of schema (T), is inconsistent. I would like to see how our naive theory of truth can be replaced by a scientifically reconstructed theory that accomplishes the same purposes without inconsistency. This new conception of truth should compare to the naive conception in much the way that relativistic notions of space and time compare to prescientific notions.

I want to restrict our attention to the problem of providing a consistent theory of truth for one's own language. To extend such an account to a more general theory, we would need to know how to translate other languages into our own. Also, I only look at sentences whose semantic features do not vary with context of utterance, thus enabling me to treat truth as a property of sentences, rather than as a relation between sentences and contexts. Understanding context dependence is an important but difficult task, and it is good methodology to tackle one hard problem at a time. Taking this methodological mandate to its extreme limits, I shall speak as if, apart from the odd behavior of the word 'true', English is a first-order language. Wildly contrary to fact, such a pretense certainly makes matters simpler.

The most straightforward way to apply Kripke's construction to develop a consistent theory of truth would be to suppose that there are truth-value gaps, and that the liar sentences are neither true nor false. Let me give three reasons, two specific and the other more general, for dissatisfaction with this approach.

The first is that, while the gap theory is intuitively very satisfying as a response to the simple liar, it fails to cope with the strengthened liar. According to the gap theory, the strengthened liar sentence is neither true nor false. *A fortiori* the sentence is not true. But that the strengthened liar sentence is not true is precisely what the strengthened liar sentence says, so that the strengthened liar sentence is true after all, contrary to the gap theory.

The second concern is closely related to the first. It is that acceptance of the gap theory renders the gap theory inexpressible. The

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

gap theorist believes that there are truth-value gaps and that the liar sentences are neither true nor false. But neither of these beliefs is, according to the gap theorist, true; they fall into the gap between truth and falsity. Even the most harmless semantic generalizations, such as 'Every true sentence is a true sentence', will be, according to the gap theorist, untrue.

If the gap theorist responds to this problem by saying that one may sincerely assert and wholeheartedly accept statements that one believes to be untrue, I shall respond that I no longer understand why the notion of truth, as he conceives it, has any importance.

If he responds that the forbidden statements seem to be untrue from the perspective of the object language, but from the perspective of the metalanguage we can see that they are true after all, I shall ask what purpose it served to go to all the trouble of developing an object-language semantics that we knew to be erroneous. Would we not be better off to follow Tarski in developing semantics entirely within the metalanguage?

The more general consideration arises from an admonition of Hartry Field[4] that our methodological standards ought not to be any lower in semantics than they are in the laboratory sciences. Now, we would not, in general, regard it as a satisfactory defense of an entrenched but inconsistent theory to weaken our logic[5] so that we hide the contradiction. For example, if we had a genetic theory to which we were particularly attached, perhaps on political grounds, and we discovered the theory to be classically inconsistent, it would be inappropriate to respond, "Classical logic does not apply when we are doing genetics."

It would be equally unsatisfactory to declare that classical logic does not apply when we are doing philosophy, and to disguise the inadequacies of our naive philosophical intuitions about truth by adopting a weakened logic in which the contradictions do not appear. Indeed, such a move would be even less appropriate in philosophy than in genetics, since in philosophy we scarcely have any empirical evidence, so that logical considerations are nearly all we have to

[4] "Tarski's Theory of Truth," this JOURNAL, LXIX, 13 (July 13, 1972): 347–375.

[5] Put less contentiously, it would not be satisfactory to restrict our methods of inference to prevent the contradiction from appearing. As I use the word 'logic', logic is the science that tells us which rules of inference preserve warranted assertibility and acceptability. On another usage, logic is an altogether more abstract science that describes relations among the truth conditions for abstract propositions, which are regarded as the bearers of truth values, rather than as the objects of belief. On the latter way of speaking, which is preferred by Kripke himself (p. 700n), truthvalueless sentences do not express propositions, so that, strictly speaking, what we say about such sentences is not a concern of logic.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

guide us toward a correct theory. If we dim the light of reason by adopting a low-wattage logic, we shall have nothing to guide us beyond our preanalytic prejudices.

We turn to the positive account, which is built upon a lesson that Kripke learned from Wittgenstein. To understand a concept properly, do not focus all your attention on how the concept behaves when it is on philosophical holiday. Look at the routine, unproblematic, nonphilosophical work the concept does. Applying this precept, we find that we use the concept of truth as a way of affirming or denying a statement without actually having to repeat the statement; it is enough to be able to name the statement. If you know what the speaker said, and I, without knowing specifically what the speaker said, know that she said something false, then we can pool our knowledge to infer the denial of the speaker's utterance. For making this sort of inference, schema (T) would be nice, but what we really need are (R1)–(R4).

As an example, let us look at a variant of Kripke's analysis of the Watergate testimony; this analysis may well turn out to be the most enduring of the "lessons of Watergate." Suppose that we have irrefutable evidence that

(1) Some of Dean's statements on May 14 are true.

and suppose further that Dean made only two statements on May 14:

(2) The dog Checkers speaks fluent Portuguese.
(3) Not everything Nixon told me yesterday is true.

Nixon's statements to Dean on May 13 were:

(4) I am not a crook.
(5) The earth moves.

From (1) and the obvious falsity of (2), we conclude that (3) is true and so either (4) or (5) is untrue. From this and the truth of (5) we conclude that (4) is untrue.

Let us see how rules (R1)–(R4) enable us to pick our way through this maze of testimony. We know from (1) that either (2) or (3) is true. Since Checkers does not speak fluent Portuguese, we know by (R3) that (2) is untrue, so (3) must be true. By (R2) we conclude that not everything Nixon told Dean was true, so that either (4) or (5) is untrue. We know that the earth moves, so that, by (R1), we know that (5) is true. Thus, we know that (4) is untrue, from which we infer, using (R4),

(6) Nixon is a crook.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(R1)–(R4) are what we want. How do we get them? Kripke shows us the way, but we need to make a subtle but crucial change in how we interpret the Kripke construction. Now we shall think of the models Kripke constructs, not as models in which sentences are true, false, or truthvalueless, but as models in which sentences are definitely true, definitely not true, or unsettled, where 'unsettled' means, not "neither true or false" but "either true or false but it is not settled which."

In allowing questions of truth to be unsettled, we are treating 'true' as a vague predicate. Now, it is not surprising that 'true' should be a vague predicate. Our naive intuitions lead us to expect that

'Harry is bald' is true iff Harry is bald.

These naive intuitions lead us to expect that, if Harry has just a little hair, so that our linguistic conventions, together with the facts about Harry's head, leave it unsettled whether Harry is bald, these facts and conventions will likewise leave it unsettled whether 'Harry is bald' is true, so that

(DB) If Harry is definitely bald, 'Harry is bald' is definitely true.
If Harry is definitely not bald, 'Harry is bald' is definitely not true.
If it is unsettled whether Harry is bald, it is likewise unsettled whether 'Harry is bald' is true.

In our naive understanding, which is governed by the acceptance of (T), 'true' is a vague predicate, but it is not simply a vague predicate like other vague predicates. Our ordinary criteria sometimes underdetermine the application of the adjective 'true', just as with other vague predicates, but they also sometimes overdetermine its application in conflicting ways. But our naive understanding of truth is inconsistent. I propose to replace our naive theory by a reconstructed theory, in which all those cases for which our ordinary understanding fails to give an univocal answer are treated alike as unsettled.

Treating 'true' as simply a vague predicate, we arrive at the analogue of (DB):

(DT) If $\ulcorner\phi\urcorner$ is definitely true, $\ulcorner\ulcorner\phi\urcorner$ is true$\urcorner$ is definitely true.
If $\ulcorner\phi\urcorner$ is definitely not true, $\ulcorner\ulcorner\phi\urcorner$ is true$\urcorner$ is definitely not true.
If $\ulcorner\phi\urcorner$ is unsettled, $\ulcorner\ulcorner\phi\urcorner$ is true$\urcorner$ is unsettled.

(DT) is precisely what Kripke's fixed-point construction gives us.

Now that we have replaced the true/false/neither trichotomy by definitely true/definitely untrue/unsettled, the 3-valued logic is no

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

longer appropriate. Instead we use supervaluations.[6] Suppose we have a first-order model A for the nonsemantical part of our language, together with an extension E for 'true', consisting in the things that definitely fall under 'true', and an anti-extension A consisting in the things that definitely fall outside 'true'. A sentence is definitely true (respectively, definitely untrue) if it comes out true (respectively, untrue) in any expansion of A to an ordinary first-order model in which the extension of 'true' is a maximal consistent set of sentences which includes E and is disjoint from A.

In accepting

(MC) The extension of 'true' is a maximal consistent set of sentences.

we are decisively renouncing our flirtation with truth-value gaps. Bivalence is a valuable feature of our ordinary conception of truth, and, since the only reason for supposing that it would not be a feature of our scientifically reconstructed conception was the futile hope that its rejection would avoid paradox, we should retain it. (MC) is what Kit Fine[7] calls a "penumbral connection," a constraint on how we may settle previously unsettled cases. A sentence is definitely true if it comes out true under any proposal for adjudicating the borderline cases which respects these constraints.

(MC) is practically useful. Imagine that, in addition to (4) and (5), Nixon had said

(7) Agnew is not a crook.

We could no longer derive (6), but we could use (MC) to derive

(8) Either Nixon or Agnew is a crook.

Here we use the fact that (MC) implies that the set of truths is closed under logical consequence. We can construct similar examples that rely upon bivalence.

(DT) gives us an explanation of why (R1)–(R4) are valid: the rules of our language permit us to assert things we are warranted in supposing to be definitely true, and (R1)–(R4) are definite-truth preserving. It also gives us a diagnosis of the pathology of the logical paradoxes. Here is a typical way of reasoning to a contradiction:

Assume that the strengthened liar sentence is true. Then 'The strengthened liar sentence is not true' is true, and so [by (R2)] the strengthened

[6] Due to Bas van Fraassen, "Singular Terms, Truth-value Gaps, and Free Logic," this JOURNAL, LXIII, 17 (September 15, 1966): 481–495.

[7] "Vagueness, Truth, and Logic," *Synthese*, XXX, 3/4 (April/May 1975): 265–300. The present account relies heavily upon Fine's work.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

> liar sentence is not true. By *reductio ad absurdum* we conclude that the strengthened liar sentence is not true, and so [by (R1)] 'The strengthened liar sentence is not true' is true. That is, we conclude that the strengthened liar sentence is true. Contradiction.

The error here, I want to claim, is in the application of (R2) within an indirect proof. Because they are definite-truth preserving, (R1)–(R4) can be legitimately applied within direct proofs that are intended to take us straightforwardly from definitely true premises to definitely true conclusions. But (R1)–(R4) cannot be legitimately applied within indirect or conditional proofs. In a similar way, the rule of necessitation in modal logic

> From $\ulcorner\phi\urcorner$ one may infer $\ulcorner\Box\phi\urcorner$.

is valid, because it preserves logical validity. But we cannot legitimately apply necessitation within indirect or conditional proofs; if we could, we would assume $\ulcorner\phi\urcorner$, necessitate, and discharge to obtain

> If ϕ, then $\Box\phi$.

We are warranted in asserting or denying $\ulcorner\phi\urcorner$ under precisely the conditions under which we are warranted in asserting or denying $\ulcorner\ulcorner\phi\urcorner$ is true$\urcorner$, but $\ulcorner\phi\urcorner$ and $\ulcorner\ulcorner\phi\urcorner$ is true$\urcorner$ do not mean the same thing, as we can see from the fact that compound sentences containing one in place of the other are not always equi-assertible. The paradoxes arise because we have correctly seen that $\ulcorner\phi\urcorner$ and $\ulcorner\ulcorner\phi\urcorner$ is true$\urcorner$ have the same assertion and denial conditions, and we have mistakenly gone on to suppose that they are intersubstitutable everywhere.

More broadly put, the paradoxes arise because we have failed to distinguish truth from definite truth.[8] This is not surprising. Vagueness is ubiquitous, but it is nearly invisible. The way we use vague terms is almost indistinguishable from the way we use precise terms; discrepancies only arise when we try to describe our usage.

Our naive concept of truth has now split into two notions, truth and definite truth. A sentence is definitely true if our linguistic conventions, together with the nonsemantic facts, insure that it is true. Definite truth is itself a vague notion. That our conventions do not insure a sentence's truth does not imply that there are conventions that forbid our making conventions that would insure its truth. So the fact that a sentence $\ulcorner\phi\urcorner$ is not definitely true does not entail that it

[8] The truth-value gap approach is a half-hearted acknowledgment of the distinction that attempts to replace a vague twofold classification by a sharp threefold classification.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

is definitely true that ⌜ϕ⌝ is not definitely true. This is important, since otherwise we could use the *Definite Liar* sentence

This sentence is not definitely true.

to reinstate the antinomy. As it is, it is perfectly possible for the definite liar to be unsettled without it being settled that the definite liar is unsettled. The theory propounded here leaves the status of the definite liar completely open.

We have seen that we can use Kripke's methods to establish that there are consistent accounts of truth which validate (R1)–(R4). This is not yet satisfactory, since we now have two semantic notions to account for—truth and definite truth—so we need to extend Kripke's methods to show that there is a consistent theory of both truth and definite truth which entails (DT) and (MC). Where Γ is a theory in the language obtained from the language of A by adding the predicates T (for 'true') and D (for 'definitely true'), let us say that ϕ is definitely true under (A, Γ) and write (A, Γ) $\models \phi$ to mean that ⌜ϕ⌝ is true in every expansion of A to a model of Γ; Γ is intended to be a system of meaning postulates that expresses everything that our linguistic conventions have determined about the extension of T and the extension of D. In this language we can formalize (DT) and (MC), so we can hope to find a theory Γ that explicitly guarantees that (DT) and (MC) are satisfied.

Our plan is to build up Γ in stages. Given Γ_β for $\beta < \alpha$, let Γ_α be the set of sentences that are made definitely true by the theory consisting in the formalized version of (DT) and (MC), together with all the sentences D(⌜ϕ⌝) and T(⌜ϕ⌝), for ϕ a member of $\bigcup_{\beta<\alpha} \Gamma_\beta$. For small values of α, we shall find that, even though the formalized version of (DT) is definitely true under (A, Γ_α), (DT) will not be satisfied by (A, Γ_α), because the extension of D will not include all the sentences that are definitely true. The situation eventually corrects itself, so that, if we set Γ_∞ equal to $\bigcup_{\alpha \in OR} \Gamma_\alpha$, we shall find that (DT) is satisfied by (A, Γ_∞) and that (A, Γ_∞) $\models \phi$ if and only if (A, Γ_∞) $\models D$(⌜ϕ⌝). To see that Γ_∞ is consistent, notice that we can expand A to a model of Γ_∞ by letting the extension of D be Γ_∞ and the extension of T be a maximal consistent set that includes Γ_∞.

This construction accomplishes less than it first appears to. It shows us that it is not unreasonable to hope to find a theory of truth which gives us (DT) and (MC); our hopes will not be dashed against a wall of contradiction. But it does not actually give us a theory that satisfies (DT) and (MC), for we have not actually written down the theory Γ_∞, and we cannot hope to do so, for the computational

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

complexity of Γ_∞ is staggering. We have merely described Γ_∞ from an external vantage point. The essentially richer metalanguage is still with us.

We can do better. One can give an effective version of this construction in which the theory Γ is actually written down, concretely and explicitly.[9] The mathematical details are somewhat intricate,[10] but their unmistakable basis is the Kripke construction.

As Tarski saw clearly, if we insist that our semantic theory entail schema (T), we shall only be able to give a semantic theory for a language if we can look at the language from the perspective of an essentially richer metalanguage. This implies that we cannot give a semantics for the very language we speak, for we have no external vantage point. On the other hand, if, as I recommend, we weaken our requirements from (T) to (DT), we find that we can give a consistent theory of truth for our own language. Thus, there will be no logical grounds for supposing that human language must lie beyond the reach of human understanding.

VANN MCGEE

University of Arizona

TRUTH, DEFINITE TRUTH, AND PARADOX*

THE STORY we are given has three parts: since (1) the paradoxes show that our "naive understanding of truth" is inconsistent, (2) a new understanding is required, which (3) is obtainable by redeployment of Saul Kripke's methods in the form of a theory which (i) does minimal violence to semantic intuition, (ii) can be formulated in the language of which it treats, (iii) discovers an error in the naive reasoning that leads to paradox, and (iv) generates no new paradoxes of its own. Just to be contrary, I quarrel with most all of this.

Suppose that our naive understanding of truth is inconsistent. Since the truth about anything, truth included, is consistent, does it not follow that naive truth theory, like any theory which misdescribes the facts, needs revision? Only, I think, if it *purports* to describe the

[9] The effective version uses A-logical consequence in place of the mathematically obstreperous supervaluational consequence.

[10] See ch. 8 of my *Truth, Vagueness, and Paradox: An Essay on the Logic of Truth* (Indianapolis: Hackett, 1989).

* Abstract of a paper to be presented in an APA symposium on Kripke's Theory of Truth, December 29, commenting on a paper by Vann McGee, this JOURNAL, this issue, 530–539.

0022-362X/89/8610/539–541 © 1989 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

facts, an assumption which sits as ill with Vann McGee's "meaning postulates" and "linguistic conventions" as with Kripke's guiding metaphor of "explaining the word 'true' to someone who does not yet understand it." Meaning something by a word involves the obligation to employ it in a certain manner, or else incorrectly; and the paradox's legendary intransigence suggests that the contradiction, if such there be, is *forced* on us by the meaning we attach to 'true'. But then its lesson is not, as the analogy with genetics suggests, that we are caught up in factual errors, but that the meaning of 'true' imposes irreconcilable obligations. If a "scientifically reconstructed" understanding of truth is indicated, that is because semantics would go better, if we meant something different by 'true'. Maybe it would. But it can hardly relieve any genuinely *philosophical* discomfort about the liar paradox, generated as it is by the meaning that we have, to be told that another is now available.

What would relieve the discomfort is an explanation of what paradox is, and how something like that can issue from procedures whose general workability has never been questioned. The "definitely" theory pins the blame on our use of a definite-truth preserving inference rule (e.g., $T^{\ulcorner}\varphi^{\urcorner}/\varphi$) in a context (conditional proof) where a (definitely) truth-preserving rule was needed; in essence, we err in supposing that, if $^{\ulcorner}\varphi^{\urcorner}$ is true, then φ. Unless it is denied, implausibly, that this is, in our usage, analytic, this diagnosis discovers an error only relative to an alien construal of 'true'. What, if anything, is the error, given the actual meaning of 'true'?

To judge by the claim that our naive understanding of truth, if paradoxical, is classically inconsistent, paradox is equated with classical inconsistency among naively acceptable classical formulae. Thus, the naively acceptable (T) $T^{\ulcorner}\varphi^{\urcorner} \leftrightarrow \varphi$, in combination with the empirical $L = \sim T^{\ulcorner}L^{\urcorner}$, classically entails $L \,\&\sim L$. By these standards, it should also be a paradox if 'Spiderman enjoys soup' lacks truth-value (in symbols, $\sim T^{\ulcorner}S^{\urcorner} \,\&\, \sim T^{\ulcorner}\sim S^{\urcorner}$); for $\sim T^{\ulcorner}S^{\urcorner} \,\&\, \sim T^{\ulcorner}\sim S^{\urcorner}$, in the presence of (T), classically entails $S \,\&\sim S$. Insistence on classical methods, in this case as in others, turns the light of reason so blindingly bright that intuitive distinctions are lost from view. The assimilation of paradox to contradiction is a case in point. That our naive understanding generates paradox *may* mean that it prescribes contradictory assertions (e.g., in the obvious notation, $O(\vdash \varphi)$ and $O(\vdash \sim\varphi)$); but it may equally mean that it prescribes incompatible actions ($O(\vdash \varphi)$ and $O(\nvdash \varphi)$); or that its instructions are incoherent ($O(\vdash \varphi)$ and $\sim O(\vdash \varphi)$), or even ill-defined ($O(\vdash \varphi$, if $\nvdash \varphi)$ and $O(\nvdash \varphi$, if $\vdash \varphi)$). On these latter hypotheses, note there is no real question of descriptive inaccuracy, because the conflicts are all at the level of obligation.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

McGee's positive proposals take off from his complaint that the gap theory, though effective against the ordinary liar ('I am false'), is powerless against the strengthened liar ('I am not true'). In fact, the same problem arises in connection with both; if 'I am false' ('I am not true') is neither true nor false, then it is *a fortiori* not false (not true), whence, given what it says, it is false (true) after all. Another problem with the gap theory is that, if true, it is inexpressible in the language of which it treats (if the liar is neither true nor false, then so is 'the liar is neither true nor false'). Likewise, though, the "definitely" theory cannot, on pain of contradiction, allow as definitely true its own (apparent) prediction that λ = 'λ is not definitely true' is not definitely true.

Perhaps this is to forget that it is not (yet) definite that λ is not definitely true (we may yet adopt conventions, given which it becomes definitely true). Such a response seems disingenuous, since any convention that would make λ definitely true would lead to contradiction. Indeed, realizing this, will we not want to enact a convention which *precludes* the determination of λ as definitely true? Only if we are willing to contradict ourselves; resolved though we may be that λ must never become definitely true, we cannot consistently express this resolution except by ascent to the metalanguage (this is the kind of thing that troubles the sleep of the ghost of Tarski's hierarchy). To make the point vivid, an appropriate formalization of 'I am not definitely true, unless everything is' seems likely to generate contradiction.

Grant that 'λ is not definitely true' is not definite; by what right do I then assert that λ is not definitely true? Perhaps the answer is that assertibility goes not with definite truth, but with truth *simpliciter;* it is true that λ is not definitely true, because λ is not entailed by facts plus current conventions. On the other hand, in calling λ indefinite, I describe it as (in McGee's words) 'either true or false, but it is not clear which', which appears to commit me to the belief that it is not clear whether λ is true. But λ is what I assert; and what business have I in asserting something such that it is not clear whether it is true? If 'λ is not definitely true' is *not* true, then it appears that 'one may sincerely assert . . . statements that one does not believe to be true'; if so, then 'I no longer understand why the notion of truth . . . has any real importance'. Could it be that we do *not* assert that λ is not definitely true? Why not, since such is evidently the case? (The ghost stirs.)

STEPHEN YABLO

University of Michigan

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

EGOISM AND MORALITY*

IS there any reason to give up rational self-interested utility maximization in favor of rule-governed behavior that enables cooperation and self-control? In Book One of the *Republic* (RU, 350b),[1] Socrates, in his main argument against Thrasymachus, attempts to demonstrate that the just man is wiser, more virtuous, stronger, and happier than the egoistic, unjust man, even if the egoist is allowed full scope and able to function as a tyrant in manipulating a society of just people for his own benefit. In the following, I shall claim that Socrates's argument is inescapable. He shows that there is an inherent difficulty in Thrasymachus's position—that anyone who aims to maximize utilities without limit (cf. G, 167)[2] can never achieve the goal of perfect satisfaction of his interests and goals (happiness) which he sets himself. I then develop this point in relation to Hobbes, David Gauthier, and Kant (cf. PP, 127, 379).[3] In doing so, I distinguish between *act egoism,* the egoism of a person who aims to maximize utilities without limit, and *rule egoism,* the egoism of a person who aims to maximize utilities through constrained cooperation with others. I argue, however, that an act egoist cannot demonstrate how it is possible within the position of act egoism to adopt the limits of rules, and thus that rule egoism is impossible.

I.

In his main skirmish with Thrasymachus, Socrates brings out one basic problem with act egoism: that it is intrinsically irrational, i.e., it can have no rule of action through which decisions can be guided concerning egoistic ends. Socrates aims to demonstrate, not only that the just man is happier than the unjust, but also that he is wiser, more virtuous, and stronger than the unjust man. Socrates's argument makes use of analogies with skilled activities. The wise or intelligent practitioner of a skill, e.g., the musician or the doctor, always has an ideal of perfection in the practice of the skill at which he aims, and which he recognizes as an absolute for himself in his practice. The clearest example of this is the case of the musician tuning a lyre. The musician aims at an ideal of perfect tuning. Once he has reached

* To be presented in an APA symposium on Egoism and Morality, December 29. Lawrence Thomas will comment; this JOURNAL, this issue, 551/2.

[1] RU = Plato's *Republic,* Paul Shorey, trans., in *The Collected Dialogues of Plato,* Edith Hamilton and Huntington Cairns, eds. (New York: Pantheon, 1961).

[2] G = David Gauthier's *Morals by Agreement* (New York: Oxford, 1986).

[3] PP = Immanuel Kant's *Perpetual Peace,* L. W. Beck, trans., in *Kant: On History* (New York: Liberal Arts, 1963).

0022-362X/89/8610/542–550 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

this ideal, he cannot better himself (RU, 349e). Different musicians will agree on the perfect tuning of the lyre, not because of convention, but because, if they are good musicians, they will all recognize the achievement of the ideal. Supposedly, moreover, two perfect physicians would agree on the diagnosis and treatment of an illness.

These examples illustrate that, in those cases where the abilities of persons in the pursuit of skilled activities can be graded absolutely, from perfect down, there must be a criterion of ideal performance of the skill against which anyone's ability can be measured. Wisdom and virtue lie in the knowledge of this ideal measure and the ability to implement it in practice. Moreover, any activity that does not proceed in relation to an ideal measure does not involve wisdom. An example of a condition of which there cannot be knowledge is "being the largest." Every size can be infinitely increased, so largeness is always relative. Because there is no "largest size," one person who constructs something large can always in principle be outdone by another who constructs something larger in relation to which the first is small.

Socrates's central point, then, is that wisdom in activities lies in knowing the ideal limit to the possibility of success in the practice of a skill. Going beyond this limit is impossible, for achieving it is perfection. Now, Socrates uses this concept of an ideal limit to argue that being just is like being wise, and being unjust, as Thrasymachus has characterized it, is like being foolish. The just man, he contends, is one who recognizes a limit to the pursuit and satisfaction of his desires both in relation to himself and in relation to others. The unjust man is one who seeks to "overreach and outdo" both the just man and the unjust (RU, 599c).

Since the range of injustice has no quantitative limit, there is no point at which the unjust man has perfected injustice. It is infinitely increasable. As a result, the unjust man must always be (a) in competition with all other men, and (b) unable to satisfy himself, because however far he has achieved injustice, there still remains an infinitely large range of greater injustice which he has not yet achieved. He can at best be only provisionally satisfied, i.e., believe that, at a given moment, he has outdone all others. But even at such a moment he must know that it is possible for him to be outdone, and the extent of his utility maximization can be infinitely increased. There is no resting place for the unjust man. He can never be secure in his achievement. His principle of maximizing utilities commits him to an unhappy life.

Thrasymachus certainly does not acknowledge such a limit. Furthermore, to the extent that utility maximization is understood as the

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

pursuit of a quantity of utility values, Thrasymachus is right, and we should always be able to imagine an increase, either in intensity or duration, or both. Indeed, experience shows that what seems the limit of, say, a pleasurable experience, can always be gone beyond. So the life of injustice must always involve dissatisfaction. Also, because the maintenance of injustice depends on natural contingency, the unjust man must be constantly in jeopardy of losing what he has. Anyone who chooses this life over the life of justice is choosing to be constantly insecure, because the goal that he sets for himself is a goal without a defined limit for maximum achievement.

Socrates, later in the *Republic,* offers a further argument to show that the unjust man cannot be happy. The second argument is based on the problem of conflicting appetites. Here Socrates compares the unjust man to a many-headed beast (RU, 588). Each appetite has its own object and strives to dominate the other appetites. There are three problems: (1) there can be no limit to the satisfaction of any appetite; (2) there can be no ideal organization of appetites which the unjust man can acknowledge; and (3) appetites change and one object seems more attractive than another from time to time. If the unjust man arbitrarily adopts some rule for giving priority to the various appetites, for example, a rule that permits satisfaction of those appetites which are compatible with the prolongation of life in a state of contentment, he then qualifies his commitment to the unjust life; for he places the resolution to obey this rule above the interests of pursuing injustice as such. He cannot, on grounds of injustice, explain why he has adopted such a rule rather than a rule of pursuing the most intense satisfaction of appetites. If he argues that injustice itself leads him to it, because it is irrational, he has granted Socrates's point.

At the very least, then, we must say that a life of injustice, the commitment to pursue utility maximization without limit, cannot lead to happiness; it must be qualified by a rule which sets limits on the pursuit of satisfaction of appetites and which gives priority to certain appetites over others. Pursuit of unlimited utility maximization is, if we are to achieve happiness, irrational.

II.

The problem with act egoism is that it cannot achieve its own end of happiness. There is no feature of it which indicates how to organize the desires and appetites. The rational egoist is doomed to dissatisfaction. If it is rational to seek happiness, then it is irrational to be an egoist.

One solution might be to show that it would be rational for the act egoist to establish an order of preferences which enables him to

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

organize and limit his various desires and appetites. No organizing rule can be derived from reflection on appetites, however. Take, for example, a rule that aims at the maintenance of maximum enjoyment over a long life, as Hobbes at times suggests (H, 86).[4] Why should anyone adopt such a rule rather than a rule that involves seeking maximum intensity of pleasure in a short life, especially since we can never know for sure what harmful contingencies life will bring? In effect, any attempt from within act egoism to resolve the problem of the inability of act egoism to achieve happiness is itself arbitrary and irrational. As a result, it is not possible to understand how a person who seeks for happiness through act egoism can accept the constraints of an ordering rule. The arbitrariness of the rule, i.e., that other rules would be equally compatible with egoism, shows that the person who adopts such a rule has no ultimate egoistic grounds for keeping the rule. Act egoism is incompatible with happiness.

In *Morals by Agreement,* Gauthier argues that it would be rational to adopt a disposition of constrained cooperation under certain conditions, and to resolve not to take advantage of "exceptions" (G, 182) when utility maximization might be furthered by doing so. Since a person with a disposition to maximize utilities without cooperating must "expect to be excluded from co-operative arrangements which he could find advantageous" (G, 183), the constrained, cooperative maximizer benefits from having opportunities that he would not otherwise have.

I suggest that there is a problem in Gauthier's account. Assuming the initial position to be act egoism, how do we get from act egoism to rule egoism? There are no adequate act-egoist grounds for supposing that it would be rational always to cooperate or that it would be rational to become a rule egoist. Furthermore, how are we to understand how the act egoist can motivate herself to abide by rules? That is, what act-egoistic motivation can there be to do this? Given the nature of act egoism itself, the act egoist can find only provisional motivation and no basis for a continuous motivation for cooperation. Gauthier suggests that the egoist can cooperate only if she adopts a *disposition* of cooperation; but how can the egoist do this on act-egoistic grounds? The adoption of such a disposition involves not taking advantages of the cooperation of others when it is possible to do so. Why should the constrained maximizer not cease cooperating just then, and then resume cooperating afterward?

These considerations reenforce the Socratic claim that we cannot resolve the problem of act egoism within the domain of act egoism

[4] H = Thomas Hobbes's *Leviathan* (New York: Liberal Arts, 1958).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

itself. We need to change standpoints, if we are to avoid the egoistic dilemma. Yet there is nothing in the standpoint of the act egoist which enables the change to take place. To argue that the change is "rational" is to suppose that there is a standpoint of objective rationality which is not part of the act-egoist position itself. A change of standpoint would involve a step away from egoism to a condition that involves rules providing limits to the satisfaction of desires and pursuit of maximizing utilities (i.e., a moral standpoint). Furthermore, it assumes a motivation capable of overriding desires and appetites.

III.

Now, the moral standpoint involves more than being constrained by rules. A moral standpoint is one that involves deliberate adherence to rules, as well as the possibility of breaking rules. We might distinguish between maximum and minimum measures to gain a moral stance. The maximum measure, if possible, would involve achieving the benefits of rule governance through elaborate cultural conditioning; but, because such a procedure would eliminate reference to egoistic ends, it would eradicate the self-reference of the egoistic entirely. Perhaps we should say that maximum-measure morality is not truly morality, because it removes the feature of personal responsibility. A minimum measure, however, would provide just the minimum conditions for resolving the problem of egoism. It must enable private egoistic acts, but do so within a framework that defines the subject as a being able to limit egoistic interests by overriding moral considerations. For minimum-measure morality, we need to add to the egoistic subject only the idea of the rational, responsible subject. The minimum measure that enables the step from act egoism to morality involves one standpoint (a) from which egoistic ends can be entertained and adopted, but (b) which acknowledges rules limiting egoistic ends within a context of self-relation and interpersonal relations. Through the minimum measure, we can preserve as much of the individualism of act egoism as possible. We can, in a modern way, characterize the standpoint of morality as that from which negotiation over principles of justice can occur.

The minimum measure enabling the step from act egoism to morality is the step whereby the agent acknowledges that he and others are persons. A person is a being capable of knowing rules and abiding by rules through recognition of their applicability. This involves the capacity to accept responsibility for right and wrong actions, and to enter into contractual relations; or, as Hobbes puts it, "a person is *he whose words or actions are considered either as his own or as representing the words or actions of an other man or any other thing to whom they are attributed*" (H, 132). The essential feature of a person is that of being an agent who understands, recognizes herself,

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

as a being with a prerogative over other beings, things (or in the case of a child or insane person, as one who is so recognized by others who represent her) (H, 46).

Although the account of the person is usually associated with Kant, Hobbes was on the brink of discovering the importance of it when he recognized that there must be a radical change in the individual's attitude to himself and others prior to entrance in a social contract. We should, however, I suggest, see the contract not as an actual agreement but as the consequence of the structure of the moral standpoint. Hobbes's position has various implications (H, 130). First is the implication that, although man is self-interested by nature, it is possible rationally to limit self-interest through the process of objective comparison with others. In this comparison, the agent must take a standpoint whereby he is not determined by self-interest, and where he is able to recognize an overriding responsibility to others which arises from recognition of the equality in status of others. A further implication is that, through this process, the agent can find a rational self that is capable of overriding his appetitive self. This rational self provides the basis for an abiding rule through which there can be morality. Furthermore, the standpoint of this rational self involves impartiality. That is, it requires that the agent recognize others as persons also and as such, equal to himself with regard to the determination of rules binding each.

Hobbes, moreover, points out one of the problems with act egoism:

> the same man in divers times differs from himself, and at one time praises—that is calls good—what another time he dispraises and calls evil (H, 131).

But he goes on to say:

> therefore so long as a man is in the condition of mere nature, which is a condition of war, private appetite is the measure of good and evil; and consequently all men agree on this: that peace is good, and therefore also the means of peace (H, 131).

Here Hobbes wrongly suggests that the problem can be resolved within the stance of act egoism. What Hobbes should have said, however, is not that all men *agree* that peace is good, but that there can be a resolution of the problem of egoism, only if there can be an agreement among men on the need for regulations governing behavior and the establishment of abiding criteria of good and evil which override the changes of appetite and temperament. This agreement can be achieved, however, only if we can make the minimum-measure change to the standpoint of the person.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

We cannot claim that this step is made *on egoistic grounds*, i.e., to secure egoistic aims more effectively. The aim of making this step is to establish a basis for one's self-relation and one's relation to others which overcomes the problem of egoism: internal conflict, and interpersonal war.

There are in effect two worlds, the world of egoism and the world of morality. Hobbes confuses the problem, however, by presenting his solution as an egoistic solution. It is not; rather, it is a solution to the dilemma of egoism, i.e., the internal inability to provide rules of behavior limiting the appetites in self-regarding and in interpersonal matters.

The limits of Hobbes's account, moreover, become evident when he argues that self-regarding virtues, such as justice of manners, gratitude, complaisance, mercy, etc., are required only to further the ends of cooperation in the civil community (H, 124–127). But there is no egoistic basis for the adoption of these virtues. They are requisites, not of egoism, but of the standpoint from which individuals regard themselves as persons. When we regard ourselves as persons, and not merely utility maximizers, we must recognize that we are members of a moral world that contains rules characterizing both the fundamental principles of interpersonal relations, and the conditions for achievement and maintenance of the status of person. Thus, they can be understood as requisites of the stance from which there can be rational control of egoistic conflicts within oneself and not merely as derived from the requisites for a peaceful coexistence.

IV.

We must now pursue further the question of whether grounds lie in the standpoint of act egoism to enable the act egoist to become a rule egoist. We must consider whether there can be constraints consistent with act egoism which enable the act egoist to become a rule egoist. Hobbes attempts to show that this is possible through the acknowledgement of the need for punishment as the constraint that enables act egoists to be governed by rules (H, 243). Laws are enforced by the constraint of others; indeed, one is obligated by a law only when the constraint of punishment motivates obedience. But no one has any obligation to submit to that constraint, if she can avoid it. Act egoism is preserved in a context that enforces rules for the egoistic advantage of mutual constraint.

The problem with Hobbes emerges when we consider that the constraint necessary to enforce laws can be supplied, only if the law not to defend another is generally obeyed. In a state of nature, however, there is no general obedience to such a law. Nor can the covenant of every man with every man create an obligation on Hobbesian grounds, unless there is constraint providing the motive for

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

fulfilling the obligation. In other words, I cannot escape from the state of nature, unless it is possible for me to be motivated to do so on some other ground than an already existing coercive force constituted by the sovereign; for no sovereign yet exists.

The problem of how to constrain the act egoist holds also for Gauthier. Although he includes an account of constraint in his account of morality (G, 6), Gauthier does not explain how constraint functions in a moral system. First of all, as noted, he does not explain how act egoism can motivate a person to be a dispositional rule egoist. Secondly, he does not supply a theory of punishment which would help to explain how anyone can be constrained by others in a way consistent with act egoism. A consequence of this is that he does not have a full account of morality, i.e., he does not explain how there can be moral blame and praise. Hence, Gauthier, like Hobbes, does not fully explain the step to a system of persons subject to laws. As part of its structure, this system requires principles of punishment which ensure that one person is subject to another and not to her own interests in the adherence to interpersonal rules.

Kant goes much further than Hobbes by suggesting that the entrance into a condition of interpersonal relations under laws does commit the individual to submit to punishment, and hence cannot be undertaken on act-egoistic grounds.[5] The condition of membership in the system of limiting rules includes submission to the law of punishment and the institutions through which it is applied. Rules of behavior of a rational being must hold independently of the private appetites, and thus be impartial and universal. A person can acknowledge these only by entering into a standpoint from which act egoism is placed secondary to obedience to rules. But, for this, a person must acknowledge the need to submit to punishment. Kant thus indicates the minimum measure that must be met for membership in the moral world.

V.

Since the resolution of the problem of act egoism depends on a motive that can lead a person to submit to rules (and punishment) without reference to egoistic advantages, rule egoism is, strictly speaking, impossible. That is, the submission to rules cannot be motivated by egoism alone. As a result, there must be a nonegoistic source of the motivation that is required for morality. This other source, I suggest, with Kant, can only be found in the thought of morality itself, and in the idea of being a person, subject to rules limiting self-regarding and interpersonal behavior, that is, the condition of morality. Indeed, as Kant contends:

[5] Kant, *The Metaphysical Elements of Justice,* John Ladd, trans. (New York: Liberal Arts, 1965), p. 105.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

> . . . our taking . . . interest in an action of duty is not suggested by an inclination, but the practical law absolutely commands it and also actually produces it.[6]

If we are to resolve the problem of egoism identified by Socrates, we must suppose, not only that we can enter a system of moral relations based on the concept of a person, but also we must be motivated to enter this system through the idea of it itself. Moreover, the consciousness of the motivation must be the ground of the consciousness that we are persons, free beings, bound by laws of the moral world.

What renders morality rational is the internal irrationality of act egoism plus the possibility of a nonegoistic, moral motivation. The consequence of this, however, is not that it is rational for the egoist to be moral, but that it is not rational to be an act egoist; and morality, with its constraints and its positions of person and impartiality, defines the standpoint of the human being with regard both to self-interest and to obligation. Within the context of persons prepared for subjection to moral rules, rules of justice can be arrived at by considerations of a contractarian sort, so long as contractarian considerations are seen only as expressions of the moral basis (with its respect for persons) from which reasoning concerning justice can proceed.

One final remark. Socrates concludes his discussion with Thrasymachus by arguing that a virtuous man is a man who has excellence in his life, i.e., a man who lives well. He then claims that living well involves well-being and therefore happiness, and concludes that the just man is happy, the unjust man unhappy (RU, 353). This has the ironical implication that justice really is the interest of the stronger, as Thrasymachus claimed. Happiness will occur, however, only if the passions and desires are so affected by the rule of behavior that they are completely satisfied or at least that the person is content. But no one can guarantee that this will happen. We can conclude only that a life governed by rules of the moral world will have a chance of happiness, indeed that the constraint of inclinations accomplished by the just man will to a great degree eliminate the pains of egoism, while an ungoverned, unjust life will have no chance of happiness at all, by virtue of the intrinsic dissatisfaction that it involves.

LESLIE MULHOLLAND

Memorial University/Newfoundland

[6] *Critique of Practical Reason*, L. W. Beck, trans. (New York: Liberal Arts, 1956), p. 84.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

DOING JUSTICE TO EGOISM*

WHAT is wrong with egoism? The answer Leslie Mulholland gives, and credits Socrates for, is that egoism does not admit of an ideal limit. This is a very provocative answer, but there is reason to be suspicious of the adequacy of its present formulation.

I.

Since Mulholland identifies egoism with immorality (of which injustice is a species)—for Thrasymachus is the egoist—the suggestion seems to be that the heart of the problem with immorality is that it constitutes a goal for which a defined limit cannot be set, as if immorality would be less of a problem if limits could be set for it. Immorality, unlike morality, overworks its adherents.

I hope that a more satisfying case than that can be made against egoism/immorality. After all, there are other maximizing goals. An altruistic person could live simply to make others better off—a goal for which a maximum limit cannot be set, and so any supposed excellence would be relative to what has been thus far achieved. Yet it seems wildly implausible to think that both this and the immoral person lead objectionable lives for essentially the same reason. This suggests that it is not open-ended maximizing as such which renders egoism/immorality unacceptable. A satisfactory account would not, as Mulholland's seems to do, lose sight of the fact that immorality characteristically harms others as well.

II.

For essentially Mulholland's reasons, I agree that David Gauthier does not offer a solution to the problem of egoism.[1] But I deny that Kantian moral theory provides us with a way of supplanting egoistic motivations. Mulholland reasons as follows: egoistic motivations cannot sustain a commitment to morality; nonegoistic motivations are necessary; and the source of nonegoistic morality is "the thought of morality itself, and the idea of being a person, subject to limiting rules on self-regarding and interpersonal behaviour." Are we being told, pace Stephen L. Darwall,[2] that, if persons sufficiently attend to moral matters, then they will in fact be moved to act morally in spite of their egoistic desires, because it is characteristic of the very nature

* Abstract of a paper to be presented in an APA symposium on Egoism and Morality, December 29, commenting on a paper by Leslie Mulholland, this JOURNAL, this issue, 542–550.

[1] See my "Rationality and Affectivity: The Metaphysics of the Moral Self," *Social Philosophy and Policy*, v (1988): 154–172.

[2] *Impartial Reason* (Ithaca: Cornell, 1983).

0022-362X/89/8610/551/2 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

of morality that its motivational content is sufficient to override egoistic desires when a person adequately attends to morality?

Although this bold claim may strike a very responsive chord in the heart of many, I do not believe that there has been an adequate defense of it. In fact, it is not even clear what would count as one.

III.

Consider the following claims: (a) egoism is an untenable theory by which to live; (b) one should be able rationally to argue the egoist into being motivated by a nonegoistic source—that is, by morality itself; (c) claim (b) is true because claim (a) is true. It is a mistake to think that, if (a) is true, then (b) and (c) must be true. For instance, although it is true that nothing touches a life like love, probably no one has or could come to believe this proposition through rational discourse. Yet this hardly counts against its truth. Sometimes the failure to convince reveals, not that the argument needs strengthening, but that the listener is not capable of appreciating it.

The problem with egoism (and the immorality born of it) is not so much that egoists are caught in the throes of maximization, but that their particular maximization endeavors make it difficult for (i) others to trust them, and so affirm their worth as individuals and (ii) for egoists to affirm others.[3] This is not true of altruists. However successful we are in life, our lives are immeasurably richer when we are affirmed by others, and when we are able to affirm others. This is a quintessential feature of our social nature. Socrates, I believe, was on to this.

It is good that we can articulate this ideal. It would be better if we could convince egoistic nonbelievers. But it would be worse if we should think our view inadequate simply because we cannot.

LAURENCE THOMAS

Syracuse University

[3] Cf. my "Ethical Egoism and Psychological Dispositions," *American Philosophical Quarterly*, XVII (1980): 73–78.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

DOMAIN EXTENSION AND THE PHILOSOPHY OF MATHEMATICS*

> Numbers are free creations of the human mind; they serve as a means for more easily and more sharply conceiving the diversity of things.
>
> Richard Dedekind

IN the philosophy of mathematics, grounds for considering specific mathematical subjects, such as group theory or complex numbers, have typically been relegated to the realm of applications. In practice, this has meant that norms governing choice of mathematical domains and theories were taken to belong to the general philosophy of science, and to arise from broadly empirical considerations. As a consequence, the possibility has not been much explored that intratheory rationality, norms concerning theory structure and organization, could have a serious role in such choices. I shall argue that they do.

I. PRAGMATICS AND MATHEMATICAL SUBJECT MATTER

In the face of a long history of vehement ontological objections, Hilbert maintained that mathematical objects such as complex numbers and points at infinity should be taken to exist, because it is mathematically fruitful to do so. In a broader context, Rudolf Carnap[1] formulates a similar view on fruitfulness more explicitly. Decisions on ontologies (in his jargon, "linguistic forms") should be determined by their "efficiency as instruments, the ratio of results achieved to the amount and complexity of the efforts required," not governed by inherently ontological considerations. W. V. O. Quine's famous reduction of ontology to what one treats as within the range of one's variables might have the same effect, as he regards the latter as subject to norms of overall scientific fruitfulness.

Unfortunately, fruitfulness seems to be a theoretical norm only in the broadest sense, somewhat like shape is a geometrical attribute in a very broad sense, but only quite special shapes lend themselves to serious geometrical study. Picking the scientifically fruitful ontology, always a matter of skill in the practitioner, might be a matter of skill

* To be presented in an APA symposium on the Philosophy of Mathematics, December 29. Philip Kitcher will comment; see this JOURNAL, this issue, 563/4. I am grateful to John Baldwin, Julius Moravscik, and members of the audience at a presentation in the Center for Philosophy of Science at the University of Pittsburgh, for their comments, encouragement, and criticisms of an earlier version.

[1] "Empiricism, Semantics, and Ontology," in *Meaning and Necessity*, 2nd. ed. (Chicago: University Press, 1956).

0022-362X/89/8610/553–562 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

only, a brute skill, if no theoretical norms apply to it. If, as Carnap and Quine seem to think, success and fruitfulness are necessarily after the scientific fact, to be enjoyed but not explained, then no philosophically interesting standards of intratheoretical rationality for acceptance of objects and theories will be forthcoming.

On the other hand, suppose we could attribute "fruitfulness-engendering" structural properties to ways of treating a subject matter, such as whether or not to use complex numbers, instead of judging the outcome of attempts to treat the subject after the fact. Then there could be reasons for success and failure: x succeeds, because x has α and having α is inherently fruitful. Then there would be a normative fact of the matter, with theoretical rather than just practical standing, whether a treatment was correct, preferable, or appropriate, based on its intrinsic merits, and regardless of attempts to carry it out.

I wish here to outline just such a theoretical justification for a variety of domain extensions from the history of mathematics. I shall show how domain extensions unify concepts, in a technical sense which covers the widely cited advantages of simplification and clarity for the use of complex numbers, projective geometry, and other examples from the mathematical and physical literature.

II. DOMAIN EXTENSION AND CONCEPTUAL UNIFICATION

The systematic adjunction of roots, or solutions to other simple conditions, as in formation of the complex numbers by adjoining imaginaries, or in adjunction of points "at infinity" in traditional geometry, may be analysed as *existential closure* and *model completion*. 'Existential closure' refers to a class of processes which attempt to round off a domain and simplify its theory by adjoining elements —more properly, it refers to the formal relationship that obtains in such a process. 'Model completion' is one of the terms employed when this process is successful. The formation of the complex numbers, and the move from affine to projective geometry, are successes of this kind. Thus, the theory of existential closure gives a theoretical basis for Hilbert's "method of ideal elements." I now sketch the theory of existential closure, to bring out when, how, and in what sense existential closure gives conceptual simplification.[2]

[2] I give a fuller account in a companion paper, in progress. For the theory of existential closure, see W. Hodges, *Building Models by Games* (Cambridge: MIT, 1985), ch. 3. For applications to algebra, see A. Macintyre, *Model Completeness,* in *Handbook of Mathematical Logic,* J. Barwise, ed. (Amsterdam: North Holland, 1977). For the model completion relationship between affine and projective geometry, see my "Interpretations and the Model Theory of the Classical Geometries," in *Models and Sets,* G. Müller and M. Richter, eds., *Springer Lecture Notes in Math 1103* (1984).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

An existential closure relation is determined by three types of items to be retained from the initial situation. These items are not necessarily obvious from the initial situation, or straightforwardly determined by it. Rather, coming to recognize a fruitful choice of preserved items is often a genuine, nontrivial intellectual achievement.

First, one initially has some individuals (in the sense of first-order logic); one wants to keep those individuals, and add some more. For example, in adjoining imaginaries, one would preserve real numbers.

Second, there are properties and relations of these individuals which one wants unaffected by the adjunction: if an individual has such a property, one wants it to retain it; if it lacks the property before the adjunction, it is to lack it afterward. Let us say these conditions are taken to be *invariant,* and denote them by Greek letters $\phi(\cdot\cdot\cdot)$, $\psi(\cdot\cdot\cdot)$, etc. In adjoining imaginaries to the real numbers, one would treat the algebraic operations as invariant. The property of being a square will not be invariant, as negatives are not squares, but will become so. Nor can the ordering of real numbers be taken to be invariant in this case, as the complex numbers cannot be made into an ordered field.

Third, one may be committed to preserve certain properties of the form $\forall \bar{u}\chi(\bar{u})$ for an invariant condition $\chi(\bar{u})$. Throughout the history of extensions of the number domain, for example, one insisted on associativity and commutativity of the "arithmetic" operations, and distributivity of multiplication over addition.[3]

Given an invariant condition $\phi(\bar{x}, \bar{y})$, we have a *solvability condition* $\exists \bar{y}\phi(\bar{x}, \bar{y})$. In algebra, for example, the $\bar{x}$ might be the coefficients of an equation, the $\bar{y}$ the unknowns. Call such individuals $\bar{y}$ *solutions* for $\bar{x}$ to the condition ϕ. Existential closure aims to add solutions, for any $\bar{x}$ where that can be done without giving up something else to which one is committed: the types of commitments just discussed entail pairwise *incompatibility* of certain solvability conditions, i.e., constraints of the form: $\sim \exists \bar{x}(\exists \bar{y}\phi(\bar{x}, \bar{y}) \wedge \exists \bar{z}\psi(\bar{x}, \bar{z}))$. Given any such commitment, we cannot adjoin a solution to the first condition for a given $\bar{x}$ once there is one to the second. For example, in arithmetic, zero has no inverse. Thus, in extending number domains, one traditionally held $(\exists y)y = 1/x$ incompatible with $x = 0$.

[3] Such preservation commitments are the content of the famous "law of the permanence of equivalent forms." But often preservation commitments are not unambiguously fixed by the subject matter prior to the adjunction. Hamilton gives up commutativity in constructing quaternions; in adjoining negatives one had to give up identities such as $\sqrt{ab} = \sqrt{a}\sqrt{b}$, arguably just as firmly established principles about quantity as the identities which were preserved.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Given such a system of commitments, one can always adjoin individuals until all solvability conditions are satisfied, except where this would violate incompatibility commitments. The resulting structures are called *existentially closed* relative to the system of commitments. Existential closure changes relations among concepts in a subject area: after existential closure, all individuals satisfy either a given solvability condition or one incompatible with it. I call this *squeezing out the middle case.*

An example: in the real numbers, one has the familiar case distinction as to whether or not a quadratic equation in one unknown is solvable. After adjoining complex numbers, the case distinction disappears: every such equation becomes solvable. The commitments subject to which complex numbers are adjoined leave solvability of quadratic equations compatible with any other (satisfiable) solvability condition.

Depending on the system of constraints (commitments), existential closure may either simplify relations among elementary-definable concepts of your theory, or render them hopelessly complex. Historical domain extensions in mathematics tend to be existential closures that simplify relations among concepts. The advantages historically cited for them may be accounted for by the theory of conceptual simplification by existential closure.

Thus, there are two kinds of systems of constraints: those which lead to fruitful existential closure, and those where existential closure gives you a mess. The difference turns on a simple *intrinsic* logical feature of systems of constraints. If every solvability condition $\exists \bar{y} \phi$ has one or more weak complements—an incompatible solvability condition $\exists \bar{z} \psi$ such that no satisfiable solvability condition is incompatible with both $\exists \bar{y} \phi$ and $\exists \bar{z} \psi$—then the squeezing out of the middle case in existential closure results in systematic conceptual unification: (a) every solvability condition becomes complementary to each of its former weak complements; (b) all weak complements of any given solvability condition become equivalent; (c) any first-order expressable condition becomes equivalent to a solvability condition —thus more complex quantifier configurations, which in most theories allow the definition of more and more complex relationships, here cannot define anything new. On the other hand, if, for a given system of constraints, some solvability condition lacks a weak complement, then squeezing out the middle case in existential closure makes that solvability condition express a condition previously only expressable by an infinitary condition; this tends to increase disasterously the complexity of your theory.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The fundamental dichotomy we encounter here provides the theoretical basis for replacing fruitfulness after the fact as a criterion of appropriateness of adjoining elements. Which behavior arises after existential closure depends on conditions attributable directly to the situation of departure—existence of weak complements. This is a global feature of systems of existential closure constraints: no single item in the system is solely responsible for weak complements or their absence.

In many familiar and successful historical adjunctions of ideal elements, additional nice things happen: every solvability condition becomes equivalent to an invariant condition from the original setting. This entails that the conceptual simplification is more radical than it would otherwise be (the invariant conditions from the original setting otherwise form a proper subset of the solvability conditions distinct after existential closure); results obtained in an existential closure are then also more readily interpreted in the original setting. In this case, existential closure is called model completion.

III. CONCEPTUAL UNIFICATION AND MATHEMATICAL UNDERSTANDING

Model completion relates an established theoretical context to a new one, intimately connected: we retain individuals and certain of their relationships and properties in the original domain. But some issues about the original context become more understandable in the model-completed one. Why?

First, a model-completed context tends to be *expressively restricted*. Only certain vocabulary from the original context remains pertinent to objects from the model-completed context. Moreover, the extension of formulas built up from this restricted vocabulary changes under domain extension, so as to cut down the range of meanings expressable to just those expressed by invariant conditions. Restricting expressiveness and unifying concepts by eliminating distinctions precludes complete solution of problems dependent on distinctions that have become nonexpressible; but it helps with problems that do not depend on such distinctions, for example, by restricting the range of choices to be considered.

Moving from real to complex algebra eliminates the possibility of expressing assertions about the ordering of real numbers. Disregarding the ordering within real algebra itself is difficult, because order remains definable from real-number multiplication. After adjunction of a root to -1, the condition previously defining the ordering becomes universally satisfied; ordering has become algebraically nonexpressible. Sets defined by polynomial inequalities ($p(x, y) < 0$) can no longer be studied in complex algebra; but algebraic geometry,

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the study of sets and relations defined by polynomial equations, is simplified: solvability conditions expressed by inequalities are no longer pertinent.

Second, suppose we are investigating a solvability condition $\exists\bar{y}\phi(\bar{x}, \bar{y})$, and we establish an implication of the form $\exists\bar{y}\phi(\bar{x}, \bar{y}) \rightarrow \forall\bar{z}\psi(\bar{x}, \bar{z})$ for some invariant condition ψ, using only inferences preserved by the domain extension. Because the connection between $\exists\bar{y}\phi(\bar{x}, \bar{y})$ and $\forall\bar{z}\psi(\bar{x}, \bar{z})$ is usually not completely tight in the original setting, we may be unable to interpret $\forall\bar{z}\psi(\bar{x}, \bar{z})$ in terms of the data expressed by $\exists\bar{y}\phi(\bar{x}, \bar{y})$. But, if the inference was not weaker than necessary, then, after model completion, $\exists\bar{y}\phi(\bar{x}, \bar{y})$ will be equivalent to $\forall\bar{z}\psi(\bar{x}, \bar{z})$, and an interpretation of $\forall\bar{z}\psi(\bar{x}, \bar{z})$ in terms of the data expressed by $\exists\bar{y}\phi(\bar{x}, \bar{y})$ is possible.

A historical example will show the effect. In the *Géométrie,* Descartes expresses relationships among line segments in a geometrical configuration by a system of several equations in the unknown segment x, and auxilliary unknowns $\bar{y}$; thus $\exists\bar{y}\bar{p}(x, \bar{y}) = 0$. Using only algebraic identities, preserved in complex algebraic geometry, to eliminate the $\bar{y}$, he obtains an equation $q(x) = 0$ in the one unknown x. But its geometrical significance is mysterious. First, its degree, algebraically the key feature. Descartes guesses that the degree determines by what means solutions may be constructed, e.g., because angle trisection problem gives an irreducible third-degree equation, it cannot be done by ruler and compass. But there is no direct way to predict the degree of its equation from the appearance of a geometrical figure. A second puzzle is the significance of the roots of q for the original geometrical configuration. The segment x in any realization of the configuration must be a root. But what about the converse?

Two centuries later, Poncelet[4] saw the geometrical picture that makes all this understandable: the roots of q in *complexified* Euclidean geometry—Euclidean geometry allowing complex coordinates—correspond exactly to the complex-Euclidean realizations of the original configuration. This is just the equivalence, after model completion, of $\exists\bar{y}\bar{p}(x, \bar{y}) = 0$ and $q(x) = 0$. There are, properly counted, as many complex-Euclidean realizations of the original configuration as the degree of q. Geometrically real roots of q may correspond to Euclidean realizations of the original configuration; or they may correspond only to realizations in which x happens to be real, but some auxilliary unknowns are complex.

Thus, the simplest and most systematic interpretation of the re-

[4] *Traité des propriétés projectives des figures* (Paris: Bachelier, 1822), sect. 53.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

sults of real analytic geometry can only be achieved in the model-completed setting. Model completions provide, not simply elegant theories of objects in esoteric idealized domains, but especially better understanding of the properties of objects and relations prior to model completion. Paradoxically, they do so by assertions, such as 'every equation of degree n has n roots', which are "false" in the original setting, though appropriate in the model completion.

IV. DOMAIN EXTENSIONS EVERYWHERE

Unification and clarification through model-completing domain extensions (or partial model completions and related constructions) are pervasive in the history of mathematics. Negative integers give unification in the theory of equations.[5] Negative occurrences of line segments allow a uniform treatment of geometric solutions to equations.[6] Similar examples could be given with adjunction of fractions, algebraic numbers, and real numbers; projective closure in geometry, differential closure in the algebraic theory of differential equations, p-adic methods (analogous to real-number methods) in contemporary number theory, and localizations of rings and modules.

The conceptual unification obtained by moving from a real to a complex setting plays an important role, not only in algebra and geometry, but also in mathematical physics. Complexification reduces the number of equivalence classes of quadratic forms, unifying the treatment of Euclidean and Minkowski metrics. The effect extends into analysis (seemingly beyond the range of the present theory), with the unification of the Schrödinger equation and the heat equation with potential; and the theory of complex-analytic continuation, which allows one to handle divergent sums in elementary particle physics. Twistor theory depends on complexification.[7]

Perhaps even more interesting and important, but not so accessible to explicit theoretical analysis, are cases in which mathematical structures or theories with model-completeness-like closure properties arose in mathematizing antecedently nonmathematical subject

[5] A transition case is G. Cardano, *The Rules of Algebra* (1545). The arrangement of the material shows that Cardano sees various cases as parallel; but, without negative coefficients, he cannot handle them uniformly.

[6] Descartes, *Geometry*, 1637. The transition is visible. In bk. I, Descartes treats geometric solution of quadratic equations, following early notes, by three different constructions, depending on signs of the coefficients; a fourth case without positive roots is not treated at all. In contrast, the uniform construction in bk. III of all roots of cubic and quartic equations is independent of signs of roots and coefficients.

[7] For a discussion of complexification in physics, see M. Eastwood, "Complexification, Twistor Theory, and Harmonic Maps from Riemann Surfaces," *Bull. American Mathematical Society*, XI (1984): 317–328. "[Complexification is] a quite precise construction. . . . many structures based on real numbers become more understandable if viewed in a complex environment" (p. 317).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

matter. Two striking cases are: traditional elementary geometry, and the natural numbers.

Of course, we cannot call traditional geometry model-complete until Tarski's first-order formalization. Nor does Tarski express the full range of concepts and concerns in the extended geometrical tradition. But the geometrical configurations and closure under geometric constructions which give the model completeness are readily recognizable in ancient and modern geometrical practice. "Idealizing" assumptions, such as the divisibility of any line segment, cannot (in a normative sense) be forced upon the skeptical by anything in physics or mathematics except the comparative simplicity and elegance of the resulting mathematical practice—these notions now receive theoretical content from the logic of domain extensions.

Elementary number theory is intractable (Gödel), and the associated first-order theories cannot be model-complete in any vocabulary adequate for number theory. This is something of a surprise, for the natural numbers would seem to be the paradigmatic example of a structure formed by "closing off" structures from prior experience —arithmetic for finite collections of numbers—by adjoining elements to satisfy solvability conditions, making addition and multiplication total operations.

The first-order theory of natural number structure is model-complete, however, in restricted vocabularies strong enough to describe that structure as arising by an existential closure process: ordering, successor, addition, 0 and 1. Only with multiplication or related notions does the theory become intractable. But in this, the natural numbers are not exceptional.[8] Moreover, number theorists have found many ways to continue the adjunction process to get tractable structures, and nonetheless obtain information about the natural numbers. The complex, real, and p-adic numbers all function in number theory as model completions of parts of the universal theory of the natural numbers.

V. EPISTEMOLOGICAL CONCLUSION

Model-completion-like domain extensions play a significant role in mathematical intellectual advance. We have some theoretical understanding of how and why. What are the epistemological consequences of this theory?

Domain extensions are moves from one theory context to another; not context-internal inferences. To accept negative quantities, one had to give up traditionally fundamental properties of quantity, such as the monotonicity of proportionalities. In moving from real to

[8] For example, with exponentiation, first-order complex-number algebra, too, is intractable.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

complex algebra, one must give up "truths," statements concerning the ordering, solvability criteria for equations, such as the discriminant formula for the quadratic equation. One must embrace different "truths," such as the solvability of every equation in one unknown. These domain extensions are truth-destroying, not deductive, truth-preserving inferences.

Nonetheless, domain extension does not start a completely new intellectual enterprise, abandoning one subject in favor of a new one, say, moving from checkers to chess, even moving from Euclidean to hyperbolic geometry. Rather, a domain-extension move brings an existing intellectual enterprise forward, realizing general formal conditions, which allow a more systematic understanding of the previous theoretical setting.

These two observations show that current logic-based epistemological traditions can hardly assign a role to domain extensions. One cannot usefully treat domain extensions as inferences within a fixed theoretical context, nor as the strictly extramathematically motivated establishment of a new theoretical context unrelated to others. The philosophical discussion of the bases of theoretical contexts has never dealt with relations between contexts. But relations between contexts are central here: we extend domains because that enhances understandibility of the original setting.

My analysis of conceptual unification by existential closure also challenges the implicit assumption of locality of intellectual virtues: that the fundamental epistemic evaluation concerns the status of individual items of information, as in 'x knows that p'. As long as we consider one fixed mathematical theory, this is plausible enough, given explicit criteria for step-by-step correctness of proofs of each individual proposition p. Once we face up to evaluating choices of theories, locality becomes doubtful. In evaluating domain extensions, locality becomes untenable. Responsibility for whether a particular existential-closure operation gives a mess or gives conceptual unification cannot be assigned to any individual item in the defining system of existential-closure constraints.

The proof-theoretic tradition in logical foundations of mathematics does evaluate certain relations between theories, by comparing proof strength or consistency strength. But domain extensions give up theorems, and have no systematic effect on consistency strength. Even when domain extensions allow one to transfer results between settings, the advantages are not proof-theoretic. Complex-number algebra is interpretable in real-number algebra, so that any proof in the complex context translates directly into the real context. But that is not the point of complex algebra.

Even when model completions have proof-theoretic advantages,

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

those do not account for the advantages of conceptual unification and enhanced systematic understandability which arise from model completion. Conceptual unification, understandability, clarity, even length of proofs, fall outside the narrow justificational concern of foundational epistemology. That concern, we should now see with the benefit of a hundred years' hindsight, fails to capture the intellectual enterprise of mathematics. Any dispassionate look at mathematical sciences should teach that the "mathematical way of knowing" seeks as much to render things understandable as it does to establish theorems or avoid error. The process of establishing deductive relations is subsidiary to the larger goal of rendering understandable.

Is a normative study of how one understands things through mathematics a reasonable philosophical goal? Some would argue that just such a study (in relation to ideal objects in general) is the very core of Plato's epistemology.[9] And in Descartes's philosophy still, clarity and distinctness of subject matter is a decisive epistemic virtue. But, from most philosophy of mathematics in this century, one would not have thought that there could or even ought to be a rigorous normative study of mathematical understanding. Matters affecting understanding are held to be irreducibly psychological, and therefore incapable of serious philosophical treatment. That inference is as blatantly fallacious as the claim that, because our grasp of the physical world necessarily involves experience, mathematics cannot possibly be the key to a theory of that world.

Much light can be shed on the topic of understandability, even by established tools from logic, though they are now not used to determine deductive strength. Mathematically characterizable differences in the organization of contexts of discourse account (at least in part) for the differences in understandability suggested by the historical record and mathematicians' judgments. Thus, there is genuine scope for epistemological inquiry into mathematical understanding. In particular, we can have explicit epistemological grounds for commitment to those domains of entities by which certain prior domains of inquiry are made more understandable.

KENNETH MANDERS

University of Pittsburgh

[9] Julius Moravcsik, "Understanding and Knowledge in Plato's Philosophy," *Neue Hefte für Philosophie,* XV/XVI (1979): 53–69.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

INNOVATION AND UNDERSTANDING IN MATHEMATICS*

MATHEMATICS students typically do not meet the complex numbers in their proper guise as ordered pairs of reals. Instead, they are informed that there is a (hitherto unmentioned) number i whose product with itself is 1. *All* numbers have square roots, *all* quadratic equations have solutions, and equations of degree n have n (not necessarily distinct) roots.

Kenneth Manders's excellent paper explores the type of transition which the extension from the reals to the complex numbers exemplifies. Manders argues for three theses.

(1) Introducing novel mathematical entities is often justified by methodological considerations internal to mathematics and independent of applications of mathematics to other areas of inquiry.
(2) Specifically, new posits increase mathematical understanding through conceptual unification.
(3) Ideas from model theory (particularly those of *existential closure* and *model completeness*) enable us to formulate precisely the conditions under which an advance in conceptual unification occurs.

I thoroughly commend (1).[1] Modifications of the language of mathematics—including the introduction, not only of terms purporting to refer to hitherto unrecognized entities, but also of new ways of referring to familiar entities—are frequently warranted by a new ability to address questions antecedently singled out as significant. Joseph Lagrange's concept of a resolvent equation helped him to understand why various tricks solve cubic and quartic equations. Newton and Leibniz relied on infinite series expansions to differentiate and integrate recalcitrant functions.

The search for understanding plays a role in the growth of mathematics and, as elsewhere in science, understanding is achieved by unifying hitherto separate domains. But (2) and (3) deserve scrutiny. In the case of (3), I am not clear why there are just two types of systems of constraints, those yielding significant conceptual unification and those introducing disastrous complications. Further problems stem from the fact that mathematicians sometimes do not value the type of conceptual unification which issues from Manders's analysis.

* Abstract of a paper to be presented in an APA symposium on the Philosophy of Mathematics, December 29, commenting on a paper by Kenneth Manders, this JOURNAL, this issue, 553–562.

[1] See *The Nature of Mathematical Knowledge* (New York: Oxford, 1983), chs. 7–9.

0022-362X/89/8610/563/4 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Rafael Bombelli introduced complex numbers to the mathematical community in the way today's students meet them. Yet complex numbers only became an accepted part of mathematical practice much later. Delay was partly due to worries about consistency which were finally resolved by geometric interpretations in the late eighteenth century. The type of unification which comes from being able to declare that all quadratic equations are solvable did not, however, strike Bombelli's contemporaries as particularly valuable: Girolamo Cardano[2] described such innovations as "subtile and useless." Only when Leibniz showed how complex factorizations could help in computing integrals and, more importantly, when Leonhard Euler connected the trigonometric and exponential functions, did it seem useful to extend the number domain. There is an important unification in Euler's $e^{i\pi} = -1$, but it is not what Manders sees occurring in cases of model completeness.

Nonetheless, I believe that Manders is right both to explore methodological principles that play a role in the growth of mathematics, and to try to formulate them precisely. I close with a larger question. Suppose this is a way in which mathematical knowledge can grow. What kinds of views of mathematical reality and mathematical progress are open to us? Can we assume that invoking entities that satisfy constraints we favor is a legitimate strategy of recognizing hitherto neglected objects that exist independently of us? From a realist perspective, the method of postulating what we want has (in Bertrand Russell's famous phrase) "the advantages of theft over honest toil." If that method is, as Richard Dedekind supposed, part of the honest trade of mathematics, is something wrong with the realist perspective?

PHILIP KITCHER

University of California/San Diego

[2] *The Great Art* (Cambridge: MIT, 1968).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

JOHNSONIAN PRIZE IN PHILOSOPHY

The editors are pleased to announce that the winner of the Johnsonian Prize in Philosophy for 1988 is Vann McGee, Assistant Professor of Philosophy, University of Arizona, for his manuscript, *Truth, Vagueness, and Paradox: An Essay on the Logic of Truth,* which will be published in November by Hackett Publishing, the co-sponsor of the prize.

Previous Johnsonian prize winners have been: 1980, Daniel Bonevac, *Reduction in the Abstract Sciences* (1981); 1981, Igal Kvart, *A Theory of Counterfactuals* (1986); 1982, Moreland Perkins, *Sensing the World* (1983); 1984, Robert J. Dunn, *The Possibility of Weakness of Will* (1986); 1985, Harry van der Linden, *Kantian Ethics and Socialism* (1988); 1986, Clyde L. Hardin, *Color for Philosophers* (1988). Also, honorable mention in 1981 went to Charles B. Guignon, *Heidegger and the Problem of Knowledge* (1983); and in 1986 to Dionysios Anapolitanos, *Leibniz: Representation, Continuity, and the Spatiotemporal.*

The Johnsonian Prize, which consists in one thousand dollars as an advance against royalty and guaranteed publication of the prize-winning book, will be offered again in 1990. The prize is designed to encourage the work of unpublished philosophers; authors are eligible who have published no book-length work. Books are selected from among manuscripts sponsored by departments of philosophy. Authors and departments who wish to submit manuscripts should observe the following guidelines:

1. Any graduate school of philosophy that confers the Ph.D. degree may submit one manuscript per year. It is not required that the work sponsored by a department be written by either a graduate or a faculty member of that department; it may, but need not be a revised version of a departmental dissertation. The closing date for the eighth prize is January 15, 1990.

2. To qualify, the manuscript must be: (a) of book length (at least 40,000 words) and in proper form for publication as a book; (b) accompanied by an abstract of up to 2000 words; (c) written by someone who has not yet published any book-length work; (d) accompanied by two professional evaluations and a letter from the sponsoring department (the editors are depending on the departments to screen out all but the most important and original work); (e) uncommitted to a publisher. Manuscripts will be returned to the authors only if accompanied by appropriate postage.

3. Preference will be given to the work of younger philosophers, especially those who have not yet been granted tenure.

Announcement of the eighth prize winner will be made in the October 1990 issue of this JOURNAL; some manuscripts may be given honorable mention. Publication of the prize-winning book will be guaranteed by the Hackett Publishing Company after the author has made what revisions are deemed necessary by the editors. The first seven prize-winning books will be on display with Hackett at the APA Division Meetings in 1989–1990.

PLEASE NOTE: The deadline for submission of manuscripts for the Johnsonian Prize for 1990 is *January 15, 1990.*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

NOTES AND NEWS

The Department of Philosophy, Logic, and Scientific Method at the London School of Economics/University of London announces the establishment of two new graduate degrees: a one-year M.S. degree in the Philosophy of Social Sciences, and an "American-style" Ph.D. program with a one-year course requirement. The M.S. in Logic and Scientific Method continues unchanged. The School would also like to announce its Lakatos Award in Philosophy of Science for an outstanding contribution to the philosophy of science in the form of a book published in English. The Award is endowed by the Latsis Foundation in memory of Imre Lakatos. To receive the Award, a successful candidate must visit the School and deliver a public lecture. Past recipients of the Award are Bas Van Fraasen and Hartry Field (1987), Michael Friedman and Philip Kitcher (1988), and Michael Redhead (1989). Information about the new graduate degrees and the Award may be obtained by writing the School, Houghton St., London WC2A 2AE, England.

The Mary Ingraham Bunting Institute of Radcliffe College announces its Fellowship Programs for 1990/91. Founded in 1960 as the Radcliffe Institute for Independent Study, the Institute is a multidisciplinary program for women scholars, scientists, artists, and writers, and is one of the major post-doctoral research centers in the country. Fellowship programs are designed to support women who wish to pursue independent work in academic and professional fields and in the arts. Office or studio space, auditing privileges, and access to libraries and other resources of Radcliffe College and Harvard University are provided. Residence in the Boston area and participation in the Institute community are required during the fellowship appointment. Fellows are expected to present their work-in-progress at public colloquia or in exhibitions. Further information may be obtained by writing the Bunting Institute of Radcliffe College, 34 Concord Ave., Cambridge, MA 02138.

Philosophy and Literature is planning a special symposium, tentatively entitled "Philosophy, Television, and Literary Culture." Questions concerning the philosophy of television cut across the boundaries of various disciplines, including sociology, politics, psychology, criticism, and aesthetics. Papers and inquiries may be directed to Denis Dutton, *Philosophy and Literature,* University of Canterbury, Christchurch, New Zealand.

The Woodrow Wilson National Fellowship Foundation announces the 1990 Charlotte W. Newcombe Doctoral Dissertation Fellowships, designed to encourage original and significant study of ethical or religious values in all fields. In addition to topics in philosophy or religion, dissertations might consider the ethical implications of foreign policy, the values determining political decisions, the moral modes of other cultures, and religious or ethical values as reflected in history or literature. Applications, which should be requested by December 1 and returned by December 15, may be obtained by writing the Newcombe Dissertation Fellowships, Woodrow Wilson National Fellowship Foundation, Box 642, Princeton, NJ 08542.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

AMERICAN PHILOSOPHICAL ASSOCIATION
EASTERN DIVISION
EIGHTY-SIXTH ANNUAL MEETING

HYATT REGENCY HOTEL, ATLANTA, GEORGIA
DECEMBER 27–30, 1989

PROGRAM

WEDNESDAY, DECEMBER 27

1:00 P.M.–10:00 P.M. REGISTRATION, *Exhibit Level*
1:00 P.M. EXECUTIVE COMMITTEE MEETING, *Board Room*

Group Meetings (see Appendix for programs)

8:00 P.M. CONFERENCE OF PHILOSOPHICAL SOCIETIES, *York*
8:00 P.M. SOCIETY FOR CLASSICAL REALISM, *Stuart*
8:00 P.M. SOCIETY FOR CHRISTIAN PHILOSOPHERS, *Essex A & B*
8:00 P.M. AMERICAN SOCIETY FOR POLITICAL AND LEGAL PHILOSOPHY, *Lancaster B & C*
8:00 P.M. WORLD PHENOMENOLOGY INSTITUTE, *Lancaster D*
8:00 P.M. LENSGRINDERS' ASSOCIATION, *Tudor*

THURSDAY, DECEMBER 28

9:00 A.M.–7:00 P.M. REGISTRATION, *Exhibit Level*
9:00 A.M.–5:00 P.M. BOOK DISPLAYS, *Ivy Hall*
9:00 A.M.–5:00 P.M. *JOURNAL OF PHILOSOPHY* DISPLAY, *Exhibit Level*

CHILD CARE arrangements. See the September or November issue of the *APA Proceedings* for further details.

PLACEMENT INFORMATION

For interviewers, 8:30 A.M.–7:00 P.M., *Brussels*
For candidates, 8:30 A.M.–7:00 P.M., *Montreal*

MORNING (9:00–11:00)

Group Meetings (see Appendix for Programs)

9:00 SOCIETY FOR CLASSICAL REALISM, *English Suite*
9:00 SOCIETY FOR THE STUDY OF PROCESS PHILOSOPHIES, *York*
9:00 INTERNATIONAL ST. THOMAS SOCIETY, *Stuart*
9:00 AMERICAN SOCIETY FOR POLITICAL AND LEGAL PHILOSOPHY *Essex A & B*
9:00 SØREN KIERKEGAARD SOCIETY, *Lancaster A*
9:00 ASSOCIATION FOR INFORMAL LOGIC AND CRITICAL THINKING *Cairo*
9:00 SOCIETY FOR PHILOSOPHY AND TECHNOLOGY, *Italian Suite*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Thursday, December 28, Morning

9:00 SOCIETY FOR SYSTEMATIC PHILOSOPHY, *French Suite*
9:00 SOCIETY FOR WOMEN IN PHILOSOPHY, *Tudor*
9:00 WORLD PHENOMENOLOGY INSTITUTE, *Austrian Suite*
9:00 INTERNATIONAL PHILOSOPHERS FOR THE PREVENTION OF NUCLEAR OMNICIDE, *Lancaster D*

LATE MORNING (11:15–1:15)

11:15 SARTRE CIRCLE, *French Suite*
11:15 SOCIETY FOR MEDIEVAL AND RENAISSANCE PHILOSOPHY *English Suite*
11:15 AMERICAN SOCIETY FOR VALUE INQUIRY, *York*
11:15 SOCIETY FOR THE ADVANCEMENT OF AMERICAN PHILOSOPHY *Lancaster B & C*
11:15 CONCERNED PHILOSOPHERS FOR PEACE, *Lancaster D*
11:15 AMERICAN ASSOCIATION FOR THE PHILOSOPHIC STUDY OF SOCIETY, *Lancaster E*
11:15 APA COMMITTEE ON BLACKS IN PHILOSOPHY, THE RADICAL PHILOSOPHY ASSOCIATION AND THE SOCIETY FOR WOMEN IN PHILOSOPHY, *Stuart*
11:15 JASPERS SOCIETY OF NORTH AMERICA, *Italian Suite*

AFTERNOON

2:00–5:00 Session I

A. Symposium: Prospects for Natural Theology, *Continental North*
Chair: Norris Clarke
Speakers: Alvin Plantinga and John Smith
Commentator: Anthony Kenny

B. Symposium: Thought of Grice, *Continental South*
Chair: Jonathan Bennett
Speakers: Judith Baker and Richard E. Grandy
Commentators: Robert Stalnaker and Richard Warner

C. Special Invited Panel on AIDS, *Phoenix*
Chair: Bernard Baumrin
Panelists: Ronald Bayer, Bernard Gert, Gunther Frankenberg, and a Representative from the Center for Disease Control

D. Colloquium: Relativism, Skepticism, and Morality, *York*
1. *Chair:* Thomas E. Hill, Jr.
Robert B. Loudon, "Through Thick and Thin: Moral Knowledge in Skeptical Times"
Commentator: Jonathan Adler

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

2. *Chair:* Michael L. Lorrado
   Michael J. Quirk, "Stout on Relativism, Liberalism, and Communitarianism"
   Commentator: James A. Montmarquet

3. *Chair:* John Riser
   Laura Westra, "Beyond MacIntyre: The Categorical Imperatives of Integrity"
   Commentator: Michael Kelly

E. Colloquium: Modern Philosophy, *Stuart*

1. *Chair:* Donald Sievert
   Lorne Falkenstein, "Berkeley's Other Argument for Other Minds"
   Commentator: Margaret Atherton

2. *Chair:* V. C. Chappell
   Mark Strasser, "Mill and Hutcheson on Evaluating Actions and Characters"
   Commentator: Jeffrey H. Reiman

3. *Chair:* Scott A. Shalkowski
   Donald P. Rutherford, "On What God Could Have Done: The Contingency of One Version of Leibniz's Principle of Sufficient Reason"
   Commentator: David Blumenfeld

F. Colloquium: Moral Psychology, *Lancaster B & C*

1. *Chair:* Nicholas Fotion
   Uma Narayan, "Varieties of Forgiveness"
   Commentator: John Deigh

2. *Chair:* Richard Capobianco
   David Schatz, "Hierarchical Theories of Freedom and the Hardening of the Hearts"
   Commentator: Eleonore Stump

3. *Chair:* Roger Sullivan
   James C. Klagge, "Virtue: Aristotle or Kant?"
   Commentator: Mark R. Herbert

G. Special Symposium Arranged by the APA Committee on Blacks in Philosophy, *French Suite*
   Topic: Afro-American Philosophy
   Chair: John Pittman
   Eric Hill, "Afro-American Philosophical Thought: Foundational Issues"
   Commentator: Stanley Browne

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Thursday, December 28, Afternoon, Session I

H. Special Panel Arranged by the APA Committee on Computer Use in Philosophy, *English Suite*

Group Meetings (see Appendix for Programs)

2:00 INTERNATIONAL SOCIETY FOR CHINESE PHILOSOPHY
Austrian Suite

2:00 AMERICAN SOCIETY FOR POLITICAL AND LEGAL PHILOSOPHY
Essex A & B

2:00 CONFERENCE OF PHILOSOPHICAL SOCIETIES, *Cairo*

2:00 NORTH AMERICAN NIETZSCHE SOCIETY, *Hong Kong*

2:00 SOCIETY FOR WOMEN IN PHILOSOPHY, *Lancaster E*

LATE AFTERNOON (5:15–7:15)

5:15 ASSOCIATION FOR PHILOSOPHY AND LIBERATION, *Italian Suite*

5:15 GHANDI-KING SOCIETY, *French Suite*

5:15 SOCIETY FOR ANCIENT GREEK PHILOSOPHY, *English Suite*

5:15 SOCIETY FOR BUSINESS ETHICS, *Stuart*

5:15 INTERNATIONAL BERKELEY SOCIETY, *Lancaster A*

5:15 AYN RAND SOCIETY, *Lancaster E*

5:15 NORTH AMERICAN KANT SOCIETY, *Austrian Suite*

5:15 AMERICAN SOCIETY FOR VALUE INQUIRY, *York*

5:15 SOCIETY FOR SYSTEMATIC PHILOSOPHY, *Lancaster D*

5:15 PERSONALIST DISCUSSION GROUP, *Flemish Suite*

5:15 SOCIETY FOR THE PHILOSOPHY OF CREATIVITY, *Spanish Suite*

5:15 SOCIETY FOR THE STUDY OF WOMEN PHILOSOPHERS, *Tudor*

5:15 CHARLES S. PEIRCE SOCIETY, *Grecian Suite*

5:15 SOCIETY FOR ASIAN AND COMPARATIVE PHILOSOPHY
Lancaster B & C

EVENING (7:30–10:30)

7:30 SOCIETY FOR PHILOSOPHY AND PUBLIC AFFAIRS, *Italian Suite*

7:30 SANTAYANA SOCIETY, *French Suite*

7:30 LEIBNIZ SOCIETY, *English Suite*

7:30 SOCIETY FOR MEDIEVAL AND RENAISSANCE PHILOSOPHY
Lancaster D

7:30 SOCIETY FOR THE STUDY OF ETHICS AND ANIMALS, *Lancaster E*

7:30 CONFERENCE OF PHILOSOPHICAL SOCIETIES, *Cairo*

7:30 RADICAL PHILOSOPHY ASSOCIATION, *Stuart*

7:30 ASSOCIATION FOR THE PHILOSOPHY OF THE UNCONSCIOUS
Spanish Suite

7:30 PHILOSOPHERS FOR SOCIAL RESPONSIBILITY, *Flemish Suite*

7:30 SOCIETY FOR PHILOSOPHERS AT WORK IN THE WORLD
Grecian Suite

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

7:30 BERTRAND RUSSELL SOCIETY, *York*
7:30 WORLD PHENOMENOLOGY INSTITUTE, *Lancaster A*
7:30 NORTH AMERICAN SOCIETY FOR SOCIAL PHILOSOPHY, *Hong Kong*

8:00 P.M.–12:00 Midnight RECEPTION, *Regency Ballroom*

FRIDAY, DECEMBER 29

9:00 A.M.–7:00 P.M. REGISTRATION, *Exhibit Level*
9:00 A.M.–5:00 P.M. BOOK DISPLAYS, *Ivy Hall*
9:00 A.M.–5:00 P.M. *JOURNAL OF PHILOSOPHY* DISPLAY, *Exhibit Level*

PLACEMENT INFORMATION

For interviewers, 8:30 A.M.–5:00 P.M., *Brussels*
For candidates, 8:30 A.M.–5:00 P.M., *Montreal*

MORNING

9:00–11:00 Session II

A. Symposium: Kripke's Theory of Truth, *Continental North*
Chair: John Burges
Speaker: Vann McGee
Commentator: Stephen Yablo

B. Invited Paper: Dilthey, Philosophy, and the Human Sciences *Essex A & B*
Chair: William Desmond
Speaker: Rudolf Makkreel
Commentator: John Scanlon

C. Invited Paper: Ethics, *Lancaster D*
Chair: Michele Moody-Adams
Speaker: Adrian Piper
Commentator: Anita Allen

D. Symposium: Analytical Marxism, *Continental South*
Chair: N. Scott Arnold
Speaker: Bill Martin
Commentator: Andrew Levine

E. Symposium: Egoism and Morality, *Tudor*
Chair: Peter L. Vallentyne
Speaker: Leslie A. Mulholland
Commentator: Laurence Thomas

F. Colloquium: Hegel, Ethical Life and the State, *York*
1. *Chair:* Andrew Ward
Andrew Buchwalter, "Hegel and the Doctrine of Expressivism"
Commentator: Willem A. De Vries

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Friday, December 29, Morning, Session II

2. *Chair:* Kenneth R. Westphal
Jeffrey Hoover, "Hegel's Critique of the Contract State"
Commentator: Michael O. Hardimon

G. Colloquium: Philosophy of Mind I, *Stuart*

1. *Chair:* Pat A. Manfredi
Thomas C. Ryckman, "Perry on Cognitive Significance"
Commentator: Leora Weitzman

2. *Chair:* Peter Ludlow
Terry Horgan and George Graham, "In Defense of Southern Fundamentalism"
Commentator: Georges Rey

H. Colloquium: Law and Explanation, *Lancaster A*

1. *Chair:* Paul K. Moser
Peter Lipton, "Contrastive Explanation"
Commentator: Jane Duran

2. *Chair:* Robert McCauley
C. Kenneth Waters, "Universal Laws of Biology"
Commentator: Samuel Mitchell

I. Colloquium: Cultural Relativism, *Lancaster B & C*

1. *Chair:* C. L. Tang
Harvey Siegel, "Feyerabend on Science and Culture"
Commentator: Alfred Nordmann

2. *Chair:* Penelope Mackie
Miriam Solomon, "On Putnam's Argument for the Inconsistency of Relativism"
Commentator: Katheryn Doran

Group Meetings (see Appendix for Programs)

9:00 SOCIETY FOR MEDIEVAL AND RENAISSANCE PHILOSOPHY
English Suite
9:00 SOCIETY FOR WOMEN IN PHILOSOPHY, *French Suite*
9:00 JASPERS SOCIETY OF NORTH AMERICA, *Italian Suite*
11:30 A.M.–1:30 P.M. BUSINESS MEETING, *Regency Ballroom*

AFTERNOON

2:15–4:15 Session III

A. The Patrick Romanell Lecture, *Continental North*
Chair: Beth Singer
Speaker: John Lachs, "Human Natures"

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

B. Invited Paper: Color, *Lancaster D*
Chair: Susan Brison
Speaker: Mark Johnston
Commentator: David Velleman

C. Invited Paper: Leibniz, *Essex A & B*
Chair: Mark Kulstad
Speaker: Robert Sleigh
Commentator: Catherine Wilson

D. Invited Paper: Legal Philosophy, *Lancaster B & C*
Chair: Alan Fuchs
Speaker: Michael Bayles
Commentator: Jules Coleman

E. Symposium: Ideal Elements and the Philosophy of Mathematics
Continental Suite
Chair: David Auerbach
Speaker: Kenneth L. Manders
Commentator: Philip Kitcher

F. Colloquium: Phenomenology, *York*
1. *Chair:* J. N. Mohanty
Richard Tieszen, "Kurt Gödel and Phenomenology"
Commentator: Steven G. Crowell

2. *Chair:* Bernard Dauenhauer
Wojciech Chojna, "Constitution of a Cultural Object"
Commentator: Mary Jane Larrabee

G. Colloquium: Philosophy of Mind II, *Stuart*
1. *Chair:* Robert Gordon
Justin Schwartz, "Davidson's Critique of Psychology"
Commentator: Louise Antony

2. *Chair:* John Heil
Don Ross, "Connectionism, GOFAI, and Psychological Explanation"
Commentator: Joseph Levine

H. Colloquim: Political Philosophy, *Tudor*
1. *Chair:* David Jones
Andrew Altman, "Critical Legal Studies versus Liberalism: Law, Deception, and Social Reality"
Commentator: Judith DeCew

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Friday, December 29, Afternoon, Session III

2. *Chair:* Ramon Lemos
David Schmidtz, "When is Original Appropriation Required?"
Commentator: George Harris

I. Colloquium: Events and Inferences, *Lancaster A*

1. *Chair:* David Coder
Susan C. Hale, "A Defense of Giving Identity Conditions for Events in Purely Spatiotemporal Terms"
Commentator: Lawrence B. Lombard

2. *Chair:* Harold Levin
Roy Sorensen, " '*P*, therefore *P*' without Circularity"
Commentator: Virginia Klenk

J. Special Symposium Arranged by the APA Committee on Pre-college Instruction in Philosophy, *English Suite*
Topic: Plato and Aristotle in the Schools
Chair: Steven S. Tigner
Panel: J. David Blankenship, John Furlong, Eugene Garver, Richard D. Parry, Thomas A. Post, and Stephen M. Rocketto

K. APA Committee on the Status of Women, *Hong Kong*
Chair: Alison Jaggar
I. Report of Committee
II. Claudia Card, "Responsibility and Moral Luck"
Commentators: Owen Flanagan and Michele Moody-Adams

Group Meetings (see Appendix for Program)

2:15 SOCIETY FOR THE PHILOSOPHIC STUDY OF GENOCIDE and THE HOLOCAUST, *French Suite*
2:15 SOCIETY FOR THE PHILOSOPHY OF SEX AND LOVE, *Lancaster E*
2:15 AMERICAN ASSOCIATION OF PHILOSOPHY TEACHERS *Flemish Suite*
2:15 RADICAL PHILOSOPHY ASSOCIATION, *Spanish Suite*
4:30 PRESIDENTIAL ADDRESS, *Regency Ballroom*
Introduction by Annette Baier
Nicholas Rescher, "Luck"

EVENING (6:30–9:30)

Group Meetings (see Appendix for Programs)

6:30 AMERICAN ASSOCIATION FOR THE PHILOSOPHIC STUDY OF SOCIETY, *Austrian Suite*
6:30 SOCIETY FOR THE PHILOSOPHIC STUDY OF GENOCIDE AND THE HOLOCAUST, *Italian Suite*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

6:30 APA COMMITTEE ON PRE-COLLEGE INSTRUCTION IN PHILOSOPHY *English Suite*
6:30 AMERICAN SOCIETY FOR VALUE INQUIRY, *Tudor*
6:30 HUME SOCIETY, *Lancaster E*
6:30 SOCIETY FOR IBERIAN AND LATIN AMERICAN THOUGHT *Flemish Suite*
6:30 SOCIETY FOR THE STUDY OF ETHICS AND ANIMALS, *Singapore*
6:30 ASSOCIATION FOR INFORMAL LOGIC AND CRITICAL THINKING *Cairo*
6:30 SOCIETY FOR SYSTEMATIC PHILOSOPHY, *Spanish Suite*
6:30 RADICAL PHILOSOPHY ASSOCIATION, *Stuart*
6:30 NORTH AMERICAN SOCIETY FOR SOCIAL PHILOSOPHY *Grecian Suite*
6:30 CONFERENCE OF PHILOSOPHICAL SOCIETIES, *York*
6:30 INTERNATIONAL SOCIETY FOR NEOPLATONIC STUDIES *Lancaster D*
6:30 G. E. MOORE SOCIETY, *Lancaster A*
6:30 SOCIETY FOR THE STUDY OF WOMEN PHILOSOPHERS, *Manila*
6:30 ASSOCIATION FOR THE PHILOSOPHY OF EDUCATION, *Hong Kong*
6:30 WORLD PHENOMENOLOGY INSTITUTE, *Essex A & B*
6:30 SOCIETY FOR THE PHILOSOPHICAL STUDY OF MARXISM and THE SOCIETY FOR THE ADVANCEMENT OF AMERICAN PHILOSOPHY *Lancaster B & C*
9:00 P.M.–12:00 Midnight RECEPTION, *Regency Ballroom*

SATURDAY, DECEMBER 30

9:00 A.M.–12:00 Noon REGISTRATION, *Exhibit Level*
9:00 A.M.–2:00 P.M. BOOK DISPLAYS, *Ivy Hall*
9:00 A.M.–12:00 P.M. *JOURNAL OF PHILOSOPHY* DISPLAY, *Exhibit Level*

PLACEMENT INFORMATION

For interviewers, 8:30 A.M.–12:00 NOON, *Brussels*
For candidates, 9:00 A.M.–12:00 NOON, *Montreal*

MORNING

9:00–11:00 Session IV

A. Invited Paper: Frege and Grammar, *Essex A & B*
Chair: Michael Resnick
Speaker: James Higginbotham
Commentator: Thomas Ricketts

B. Invited Paper: Plato and Aristotle on Women, *Lancaster B & C*
Chair: Martha Nussbaum
Speaker: Elizabeth Spelman
Commentator: Marcia Homiak

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Saturday, December 30, Morning, Session IV

C. Book Discussion: *The Shaky Game* by Arthur Fine
Continental Room

Chair: Robert Burton
Discussant: R. I. G. Hughes
Respondent: Arthur Fine

D. Symposium: "Self-Interest" in Adam Smith, *Lancaster A*
Chair: Jerome B. Schneewind
Speaker: Patricia H. Werhane
Commentator: Charles L. Griswold, Jr.

E. Symposium: Impartiality and the Moral Point of View, *Tudor*
Chair: Lawrence Becker
Speaker: Marilyn Friedman
Commentator: Howard McGary

F. Colloquium: Derrida and the Philosophical Canon
Continental South

1. *Chair:* Leonard Lawlor
J. Claude Evans, "What Is Pragrammatology? A Case Study"
Commentator: Jacob Owensby

2. *Chair:* Ronald Hall
Michael Murray, "A Theory of Canonization"
Commentator: Hugh Silverman

G. Colloquium: Medieval Philosophy, *Lancaster D*

1. *Chair:* Jude Dougherty
Pamela Hall, "Aristotelian Practical Wisdom and Thomistic Natural Law"
Commentator: William E. Mann

2. *Chair:* Mary T. Clark
Mark P. Drost, "Intentionality in Aquinas's Theory of the Emotions"
Commentator: Scott MacDonald

H. Colloquium: Imagination and Pain, *York*

1. *Chair:* Gene James
Richard Garner, "Nagel on the Badness of Pain"
Commentator: Henry West

2. *Chair:* William Maker
A. Chadwick Ray, "Imagination in Practical Reason"
Commentator: Michael J. Zimmerman

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

I. Colloquium: Carnap and Wittgenstein, *Stuart*
1. *Chair:* Leonard Carrier
Alan Richardson, "Idealisms in the *Aufbau*"
Commentator: Robert Burch

2. *Chair:* Ned Markosian
Donna Summerfield, "*Philosophical Investigations* 201: 'The Answer Was . . .'"
Commentator: Edward H. Minar

J. Special Symposium Arranged by the APA Committee on Medicine, *Hong Kong*
Topic: Quality of Life: Can It Be Measured?
Chair: Bernard Baumrin
Speaker: Alex C. Michalos
Commentator: Nicholas Rescher

K. Special Symposium Arranged by the APA Committee on the Teaching of Philosophy, *Cairo*
Topic: The Undergraduate Major in Philosophy
Panel: Robert Audi, Gary Iseminger, Anita Silvers, Laurence Thomas, and Merold Westphal

11:15 A.M.–12:15 P.M. BUSINESS MEETING (Second Session, if needed)
Regency Ballroom

AFTERNOON

1:30–4:30 Session V

A. Symposium: Thought of Wittgenstein, *Continental North*
Chair: David Sachs
Speakers: Warren Goldfarb and Crispin Wright
Commentator: John McDowell

B. Book Discussion: *The Persistence of Reality* by Joseph Margolis
Continental South
Chair: Don Ihde
Discussants: Ralph Sleeper, Ernest Sosa, and Marx Wartofsky
Respondent: Joseph Margolis

C. Symposium: Foucault, *Essex A & B*
Chair: Thomas Flynn
Speakers: Ladelle McWhorter, Judith Butler, and David F. Gruber

D. Colloquium: Ethics of Nuclear Deterrence, *Lancaster B & C*
1. *Chair:* Barry L. Gan
Walter Sinnott-Armstrong, "The Wrongful Intentions Principle: A Response to Lango"
Commentator: John W. Lango

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Saturday, December 30, Afternoon, Session V

2. *Chair:* Ofelia Schutte
David A. Hoekema, "Is Strategic Defense Morally Defensible?"
Commentator: Daniel M. Farrell

3. *Chair:* George Hampsch
Stephen Nathanson, "Kavka on Kennedy and the Cuban Missile Crisis"
Commentator: Steven P. Lee

E. Colloquium: Ancient Philosophy, *Lancaster A*

1. *Chair:* David H. Calhoun
Herbert Granger, "The Nature of Aristotle's Psychology"
Commentator: L. A. Kosman

2. *Chair:* John Anton
Victor Caston, "The Stoics on Concepts"
Commentator: Wolfgang Mann

3. *Chair:* William Edwards
James T. H. Martin, "Aristotle on the Separateness on Mind"
Commentator: Robin Smith

F. Colloquium: Morality and the Emotions, *Tudor*

1. *Chair:* Daniel E. Wueste
Edmund Abegg, "Caring and the Consequentialist Escape from Particularism"
Commentator: Mary B. Mahowald

2. *Chair:* Heidi Malm
William Neblett, "Rights and Feelings: A Study of Rights to Possess, and Rights to Express, Feelings"
Commentator: Ferdinand Schoeman

3. *Chair:* Harry van der Linden
Glenn A. Hartz, "Desire and Emotion in the Virtue Tradition"
Commentator: Louis Pojman

1:30 EXECUTIVE COMMITTEE MEETING, *Board Room*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

PROGRAM PARTICIPANTS
EIGHTY-SIXTH ANNUAL MEETING

ABEGG, Edmund, University of Pennsylvania/Edinboro V-F
ADLER, Jonathan, City University of New York/Brooklyn College I-D
ALLEN, Anita, Georgetown University II-C
ALTMAN, Andrew, George Washington University III-H
ANTON, John, University of South Florida V-E
ANTONY, Louise, North Carolina State University III-G
ARNOLD, N. Scott, University of Alabama/Birmingham II-D
ATHERTON, Margaret, University of Wisconsin/Milwaukee I-E
AUDI, Robert, University of Nebraska/Lincoln IV-K
AUERBACH, David, North Carolina State University III-E
BAIER, Annette, University of Pittsburgh
Introduction: Presidential Address
BAKER, Judith, York University I-B
BAUMRIN, Bernard, City University of New York/Lehman College I-C, IV-J
BAYER, Ronald, Columbia University I-C
BAYLES, Michael, Florida State University III-D
BECKER, Lawrence, College of William & Mary, Virginia IV-E
BENNETT, Jonathan, Syracuse University I-B
BLANKENSHIP, J. David, State University of New York/New Paltz III-J
BLUMENFELD, David, Georgia State University I-E
BRISON, Susan, New York University III-B
BROWNE, Stanley, Youngstown State University and Morgan State University I-G
BUCHWALTER, Andrew, University of Northern Florida II-F
BURCH, Robert, Texas A & M University IV-I
BURGESS, John, Princeton University II-A
BURTON, Robert, University of Georgia IV-C
BUTLER, Judith, Johns Hopkins University V-C
CALHOUN, David H., Gonzaga University V-E
CAPOBIANCO, Richard, Stonehill College, Massachusetts I-F
CARD, Claudia, University of Wisconsin/Madison III-K
CARRIER, Leonard, University of Miami IV-I
CASTON, Victor, University of Texas/Austin V-E
CHAPPELL, V. C., University of Massachusetts/Amherst I-E
CHOJNA, Wojciech, LaSalle University III-F
CLARK, Mary T., Manhattanville College, New York IV-G
CLARKE, Norris, Fordham University I-A
CODER, David, National Endowment for the Humanities III-I
COLEMAN, Jules, Yale University III-D
CROWELL, Steven G., Rice University III-F
DAUENHAUER, Bernard, University of Georgia III-F
DECEW, Judith, Clark University III-H
DEIGH, John, Northwestern University I-F

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

DESMOND, William, Loyola University/Baltimore II-B
DEVRIES, Willem A., University of New Hampshire II-F
DORAN, Katheryn, Wellesley College, Massachusetts II-I
DOUGHERTY, Jude, Catholic University of America IV-G
DROST, Mark P., University of Rochester IV-G
DURAN, Jane, University of California/Santa Barbara II-H
EDWARDS, William, Emory University V-E
EVANS, J. Claude, Washington University IV-F
FALKENSTEIN, Lorne, University of Western Ontario, Canada I-E
FARRELL, Daniel M., Ohio State University V-D
FINE, Arthur, Northwestern University IV-C
FLANAGAN, Owen, Wellesley College, Massachusetts III-K
FLYNN, Thomas, Emory University V-C
FOTION, Nicholas, Emory University I-F
FRANKENBERG, Gunther, Frankfurt, Germany I-C
FRIEDMAN, Marilyn, Purdue University IV-E
FUCHS, Alan, College of William & Mary, Virginia III-D
FURLONG, John, Coppin State College, Maryland III-J
GAN, Barry L., St. Bonaventure University V-D
GARNER, Richard, Ohio State University IV-H
GARVER, Eugene, St. John's University III-J
GERT, Bernard, Dartmouth College, New Hampshire I-C
GOLDFARB, Warren, Harvard University V-A
GORDON, Robert, University of Missouri/St. Louis III-G
GRAHAM, George, University of Alabama/Birmingham II-G
GRANDY, Richard E., Rice University I-B
GRANGER, Herbert, Wayne State University V-E
GRISWOLD, Charles, Howard University IV-D
GRUBER, David F., Northeast Missouri State University V-C
HALE, Susan C., Texas A & M University III-I
HALL, Pamela, Emory University IV-G
HALL, Ronald, Francis Marion College, South Carolina IV-F
HAMPSCH, George, Holy Cross College, Massachusetts V-D
HARDIMON, Michael O., Massachusetts Institute of Technology II-F
HARRIS, George, College of William & Mary, Virginia III-H
HARTZ, Glenn A., Ohio State University/Mansfield V-F
HEBERT, Mark R., University of North Carolina/Chapel Hill I-F
HEIL, John, Davidson College, North Carolina III-G
HIGGINBOTHAM, James, Massachusetts Institute of Technology IV-A
HILL, Eric, University of Louisville I-G
HILL, JR., Thomas E., University of North Carolina/Chapel Hill I-D
HOEKEMA, David A., University of Delaware V-D
HOMIAK, Marcia, Occidental College, California IV-B
HOOVER, Jeffrey, Coe College, Iowa II-F
HORGAN, Terry, Memphis State University II-G
HUGHES, R. I. G., University of South Carolina/Columbia IV-C
IHDE, Don, State University of New York/Stony Brook V-B

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ISEMINGER, Gary, Charleton College, Minnesota IV-K
JAGGAR, Alison, University of Cincinnati III-K
JAMES, Gene, Memphis State University IV-H
JOHNSTON, Mark, Princeton University III-B
JONES, David, College of William & Mary, Virginia III-H
KELLY, Michael, Columbia University I-D
KENNY, Anthony, Oxford University/Balliol College I-A
KITCHER, Philip, University of California/San Diego III-E
KLAGGE, James C., Virginia Polytechnic Institute & State University I-F
KLENK, Virginia, University of West Virginia III-I
KOSMAN, L. A., Haverford College, Pennsylvania V-E
KULSTAD, Mark, Rice University III-C
LACHS, John, Vanderbilt University III-A
LANGO, John W., City University of New York/Hunter College V-D
LARRABEE, Mary Jane, DePaul University III-F
LAWLOR, Leonard, Memphis State University IV-F
LEE, Steven P., Hobart & William Smith College, New York V-D
LEMOS, Ramon, University of Miami III-H
LEVIN, Harold, North Carolina State University III-I
LEVINE, Andrew, University of Wisconsin/Madison II-D
LEVINE, Joseph, North Carolina State University III-G
LIPTON, Peter, Williams College, Massachusetts II-H
LOMBARD, Lawrence B., Wayne State University III-I
LORRADO, Michael L., University of North Carolina/Chapel Hill I-D
LOUDON, Robert B., University of Southern Maine I-D
LUDLOW, Peter, State University of New York/Stony Brook II-G
MacDONALD, Scott, University of Iowa IV-G
MACKIE, Penelope, Virginia Commonwealth University II-I
MAHOWALD, Mary B., Case Western Reserve University V-F
MAKER, William, Clemson University IV-H
MAKKREEL, Rudolf, Emory University II-B
MALM, Heidi, University of Nebraska V-F
MANDERS, Kenneth L., University of Pittsburgh III-E
MANFREDI, Pat A., Hamilton College, New York II-G
MANN, William E., University of Vermont IV-G
MANN, Wolfgang, Middlebury College, Connecticut V-E
MARGOLIS, Joseph, Temple University V-B
MARKOSIAN, Ned, Lawrence University IV-I
MARTIN, Bill, University of Kansas II-D
MARTIN, James T. H., St. John's University V-E
McCAULEY, Robert, Emory University II-H
McDOWELL, John, University of Pittsburgh V-A
McGARY, Howard, Rutgers University/New Brunswick IV-E
McGEE, Vann, University of Arizona II-A
McHORTER, Ladelle, Northeast Missouri State University V-C
MICHALOS, Alex C, University of Guelph, Canada IV-J

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

MINAR, Edward H., Johns Hopkins University IV-I
MITCHELL, Samuel, Mt. Holyoke College, Massachusetts II-H
MOHANTY, J. N., Temple University III-F
MONTMARQUET, James A., Nashville, Tennessee I-D
MOODY-ADAMS, Michele, University of Rochester II-C, III-K
MOSER, Paul K., Loyola University of Chicago II-H
MULHOLLAND, Leslie A., Memorial University of Newfoundland, Canada II-E
MURRAY, Michael, Vassar College, New York IV-F
NARAYAN, Uma, Rutgers University/New Brunswick I-F
NATHANSON, Stephen, Northwestern University V-D
NEBLETT, William, Occidental College, California V-F
NORDMANN, Alfred, University of South Carolina/Columbia II-I
NUSSBAUM, Martha, Brown University IV-B
OWENSBY, Jacob, Jacksonville University IV-F
PARRY, Richard D., Agnes Scott College, Georgia III-J
PIPER, Adrian, University of California/La Jolla II-C
PITTMAN, John, City University of New York/Graduate Center I-G
PLANTINGA, Alvin, University of Notre Dame I-A
POJMAN, Louis, University of Mississippi V-F
POST, Thomas A., Millbrook School, New York III-J
QUIRK, Michael J., Bellerose, New York I-D
RAY, A. Chadwick, Central College, Iowa IV-H
REIMAN, Jeffrey H., American University I-E
RESCHER, Nicholas, University of Pittsburgh IV-J
Presidential Address
RESNICK, Michael, University of North Carolina IV-A
REY, Georges, University of Maryland II-G
RICHARDSON, Alan, Philadelphia, Pennsylvania IV-I
RICKETTS, Thomas, University of Pittsburgh IV-A
RISER, John, University of Central Florida I-D
ROCKETTO, Stephen M., Grasso-Southern Regional Vocational Technical School, Connecticut III-J
ROSS, Don, University of Western Ontario, Canada III-G
RUTHERFORD, Donald P., Emory University I-E
RYCKMAN, Thomas C., Lawrence University II-G
SACHS, David, Johns Hopkins University V-A
SCANLON, John, Duquesne University II-B
SCHATZ, David, Flushing, New York I-F
SCHMIDTZ, David, Yale University III-H
SCHNEEWIND, Jerome B., Johns Hopkins University IV-D
SCHOEMAN, Ferdinand, University of South Carolina/Columbia V-F
SCHUTTE, Ofelia, University of Florida V-D
SCHWARTZ, Justin, Ohio State University III-G
SHALKOWSKI, Scott A., University of Oklahoma I-E
SIEGEL, Harvey, University of Miami II-I
SIEVERT, Donald, University of Missouri I-E

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

SILVERMAN, Hugh, State University of New York/Stony Brook IV-F
SILVERS, Anita, San Francisco State University IV-K
SINGER, Beth, City University of New York/Brooklyn College III-A
SINNOTT-ARMSTRONG, Walter, Dartmouth College, New Hampshire V-D
SLEEPER, Ralph, City University of New York/Queens College V-B
SLEIGH, Robert, University of Massachusetts/Amherst III-C
SMITH, John, Yale University I-A
SMITH, Robin, Kansas State University V-E
SOLOMON, Miriam, University of Cincinnati II-I
SORENSEN, Roy, New York University III-I
SOSA, Ernest, Brown University V-B
SPELMAN, Elizabeth, Smith College, Massachusetts IV-B
STALNAKER, Robert, Massachusetts Institute of Technology I-B
STRASSER, Mark, Washington University I-E
STUMP, Eleonore, Virginia Polytechnic Institute & State University I-F
SULLIVAN, Roger, University of South Carolina/Columbia I-F
SUMMERFIELD, Donna, Hamilton College, New York IV-I
TANG, C. L., California State University/Long Beach II-I
THOMAS, Laurence, Syracuse University II-E, IV-K
TIESZEN, Richard, San Jose State University III-F
TIGNER, Steven S., University of Toledo III-J
VALLENTYNE, Peter L., University of Western Ontario, Canada II-E
VAN DER LINDEN, Harry, University of North Carolina/Chapel Hill V-F
VELLEMAN, David, University of Michigan III-B
WARD, Andrew, Saint Olaf College, Minnesota II-F
WARNER, Richard, University of Southern California I-B
WARTOFSKY, Marx, City University of New York/Baruch College V-B
WATERS, C. Kenneth, Rice University II-H
WEITZMAN, Leora, University of Wisconsin/Madison II-G
WERHANE, Patricia H., Loyola University of Chicago IV-D
WEST, Henry, MacAlester College, Minnesota IV-H
WESTPHAL, Kenneth R., University of New Hampshire II-F
WESTPHAL, Marold, Fordham University IV-K
WESTRA, Laura, University of Toledo I-D
WILSON, Catherine, University of Oregon III-C
WRIGHT, Crispin, University of Michigan V-A
WUESTE, Daniel E., Bowling Green State University V-F
YABLO, Stephen, University of Michigan/Ann Arbor II-A
ZIMMERMAN, Michael J., University of North Carolina/Greensboro IV-H

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

APPENDIX: PROGRAMS OF GROUP MEETINGS

WEDNESDAY, DECEMBER 27

I. EVENING (8:00–11:00)

1. 8:00 CONFERENCE OF PHILOSOPHICAL SOCIETIES, *York*
   Topic: Annual Business Meeting

2. 8:00 SOCIETY FOR CLASSICAL REALISM, *Stuart*
   Topic: Charles S. Peirce and Realism
   Chair: Raymond L. Langley
   I. Douglas R. Anderson, "Idealism and Realism in Peirce"
   II. Patrick Sullivan, "C.S. Peirce and the Current Debate over Scientific Realism"

3. 8:00 SOCIETY OF CHRISTIAN PHILOSOPHERS, *Essex A & B*
   Topic: The Significance of Narrative for Theology
   Chair: Robert M. Adams
   George Lindbeck, "The Significance of Narrative for Theology"
   Commentators: William Alston and Philip Quinn

4. 8:00 AMERICAN SOCIETY FOR POLITICAL AND LEGAL PHILOSOPHY *Lancaster B & C*
   Topic: Virtue
   George Sher, "An Epistemology of the Virtues"

5. 8:00 WORLD PHENOMENOLOGY INSTITUTE, *Lancaster D*
   Topic: S. Laycock's *Foundations for a Phenomenological Theology* and A-T. Tymieniecka's *The Three Movements of the Soul* (Book 2)
   Chair: Walter Law
   I. Alan Olson, "Two Approaches to the Divine in Laycock and Tymieniecka"
   II. Paul Brockelmans, "Different Ways to the Top?"
   Commentator: Krystyna Gorniak-Kocilkowska

6. 8:00 LENSGRINDERS' ASSOCIATION, *Tudor*
   Reception for Job Seekers

THURSDAY, DECEMBER 28

II. MORNING (9:00–11:00)

1. 9:00 SOCIETY FOR CLASSICAL REALISM, *English Suite*
   Topic: Charles S. Peirce and Realism
   Chair: Fritz Wenisch
   I. Mary Carmen Rose, "Peirce, Dewey, and the Future of Philosophy"
   II. Carl R. Hausman, "Peirce's Evolutionary Realism"

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

2. 9:00 SOCIETY FOR THE STUDY OF PROCESS PHILOSOPHIES, *York*
Topic: Anticipations of Cultures
Chair: Joseph Grange
Speakers: Roger T. Ames and David L. Hall

3. 9:00 INTERNATIONAL ST. THOMAS SOCIETY, *Stuart*
Chair: Mary Rose Barral
Robert Spitzer, "Toward a Thomistic Metaphysics of Time"

4. 9:00 AMERICAN SOCIETY FOR POLITICAL AND LEGAL PHILOSOPHY
Essex A & B
Topic: Virtue
Ronald Beiner, "The Moral Vocabulary of Liberalism"
Business Meeting

5. 9:00 SØREN KIERKEGAARD SOCIETY, *Lancaster A*
Chair: John Donnelly
I. Eric J. Ziolkowski, "The Concept of Aesthetic: A Semantic Leap from Baumgarten to Kierkegaard"
Commentator: George Connell
II. Benjamin Daise, "The Will to Truth"
Commentator: Louis Pojman
III. Stephen C. Rowntree, "Self-Love versus Self-Hate: The Triumph of the Kierkegaardian Therapeutic"
Commentator: C. Stephen Evans

6. 9:00 ASSOCIATION FOR INFORMAL LOGIC AND CRITICAL THINKING
Cairo
I. Presidential Address
Chair: John Hauglund
Mark Weinstein, "A Research Agenda for Informal Logic and Critical Thinking"
II. *Topic:* Matthew Lipman's Program for Philosophy for Children: A Critical Appreciation (Co-sponsored with the APA Committee on Pre-college Philosophy)
Chair: Gareth Matthews
Speakers: Louis Katzner, James S. Katzner, and Harvey Siegel
Respondent: Matthew Lipman

7. 9:00 SOCIETY FOR PHILOSOPHY AND TECHNOLOGY, *Italian Suite*
Topic: Pragmatism and Social Criticism in a Technological Age
Chair: Paul T. Durbin
Speakers: Thomas Alexander and Larry Hickman
Commentator: Richard J. Bernstein

8. 9:00 SOCIETY FOR SYSTEMATIC PHILOSOPHY, *French Suite*
I. Kenley R. Dove, "Individuality in the Modern World"
II. Edward Halper, "On Individuality"

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Thursday, December 28, Morning, II

9. 9:00 SOCIETY FOR WOMEN IN PHILOSOPHY, *Tudor*
Topic: Gender and Science
Chair: Jana Sawicki
I. Elizabeth Potter, "Gender Politics in the Origins of Modern Science"
II. Helen Longino and Evelyn Hammonds, "Feminism and the Hard Sciences: The Case of Physics"

10. 9:00 WORLD PHENOMENOLOGY INSTITUTE, *Austrian Suite*
Chair: Edward Farley
I. Jude Dougherty, "The Origin of Religion"
II. Robert R. Williams, "Some Reflections on Laycock's *Phenomenological Foundations of Religion*"
Commentator: Marlies Kronegger

11. 9:00 INTERNATIONAL PHILOSOPHERS FOR THE PREVENTION OF NUCLEAR OMNICIDE, *Lancaster D*
Topic: Does Ideology Have a Place in the Struggle for World Peace?
Chair: George Hampsch
I. Erich Hahn, "The Challenge for Ideology in the Securing of Peace"
II. Ofelia Schutte, "On the Relation between Ideology and World Peace: Some Reflections"
III. Michael Howard, "Identity and Difference in the Struggle for Peace"

III. LATE MORNING (11:15–1:15)

1. 11:15 SARTRE CIRCLE, *French Suite*
Chair: Phyllis S. Morris
I. David Detmer, "What Is Objectivity? Sartre versus the Journalists"
Commentator: Thomas R. Flynn
II. Linda Bell, "Sartrean Ethics and Feminism"
Commentator: Julien S. Murphy

2. 11:15 SOCIETY FOR MEDIEVAL AND RENAISSANCE PHILOSOPHY *English Suite*
Executive Committee Session

3. 11:15 AMERICAN SOCIETY FOR VALUE INQUIRY, *York*
Topic: Rights and Duties
Chair: Thomas Magnell
I. Sander Lee, "A Contemporary Debate on the Proper Role of Rights" (Presidential Address)
II. Business Meeting

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

4. 11:15 SOCIETY FOR THE ADVANCEMENT OF AMERICAN PHILOSOPHY *Lancaster B & C*

Topic: Dewey's Epistemology Today
Chair: William J. Gavin
I. Hilary Putnam and Ruth Anna Putnam, "Epistemology As Hypothesis"
Commentator: Ralph Sleeper
II. H. S. Thayer, "Dewey and the Theory of Knowledge"
Commentator: Larry Hickman

5. 11:15 CONCERNED PHILOSOPHERS FOR PEACE, *Lancaster D*

I. *Topic:* Symposium on Robert L. Holmes's *On War and Morality*
Chair: Barry L. Gan
Speakers: Robert L. Holmes, Jeffrie Murphy, and James Sterba
II. Business Meeting

6. 11:15 AMERICAN ASSOCIATION FOR THE PHILOSOPHIC STUDY OF SOCIETY *Lancaster E*

Topic: Jan Narveson's *The Libertarian Idea*
Chair: Tibor R. Machan
Speakers: Peter A. Danielson, Jeremy Shearmur, and Stuart Warner
Respondent: Jan Narveson

7. 11:15 APA COMMITTEE ON BLACKS IN PHILOSOPHY, THE RADICAL PHILOSOPHY ASSOCIATION, and THE SOCIETY FOR WOMEN IN PHILOSOPHY, *Stuart*

Topic: Affirmative Action and Social Justice
Chair: Carolyn Magid
Speakers: Shirley Clark Franklin (City Solicitor/Atlanta), Gertrude Ezorsky, and Iris Young

8. 11:15 JASPERS SOCIETY OF NORTH AMERICA, *Italian Suite*

Chair: Harold H. Oliver
I. Krystyna Groniak-Kocilkowska, "Individual and Society in Jaspers and Marx"
Commentator: Greg Walters
II. William Kluback, "Philosophy's Discordant Voice"
Commentator: Alan Olson

IV. AFTERNOON (2:00–5:00)

1. 2:00 INTERNATIONAL SOCIETY FOR CHINESE PHILOSOPHY *Austrian Suite*

Chair and Commentator: Kwang-Sae Lee
I. Thomas L. Benson, "Evil and Theodicy in Early Confucian Thought"
II. Joel R. Smith, "Confucianism and MacIntyre on Virtue"
III. Jay G. Williams, "Plato and I Ching"

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Thursday, December 28, Afternoon, IV

2. 2:00 AMERICAN SOCIETY FOR POLITICAL AND LEGAL PHILOSOPHY *Essex A & B*

Topic: Virtue
David Luban, "Virtues and Vices of Judges"

3. 2:00 CONFERENCE OF PHILOSOPHICAL SOCIETIES, *Cairo*
Topic: Workshop on Philosophy and Film
Chair: Charles Hornbeck
I. John A. Loughney, "Values and Social Thought in Film"
II. Sander Lee, "Existential Themes in Hitchcock's *Vertigo*"
III. Charles List, "Can Films Be Forged?"
IV. David Ward, "Themes in Philosophy and Film"
V. Business Meeting

4. 2:00 NORTH AMERICAN NIETZSCHE SOCIETY, *Hong Kong*
Topic: Symposium: Nietzsche's Madness, 1889–1989
Chair: Gary Shapiro
Speakers: David Allison, Daniel Breazeale, David Farrell Krell, and Alphonso Lingis

5. 2:00 SOCIETY FOR WOMEN IN PHILOSOPHY, *Lancaster E*
I. In Honor of Distinguished Women Philosophers
Chair: Margaret Walsh
Honoree: Hazel E. Barnes
Hazel Barnes, "Sartre's Sexism: Essential or Contingent"
Commentators: Linda Bell and Judith Butler
II. Business Meeting
Chair: Jana Sawicki

V. EARLY EVENING (5:15–7:15)

1. 5:15 ASSOCIATION FOR PHILOSOPHY AND LIBERATION, *Italian Suite*
Chair: Mary Christine Morkovsky
Robert Gibbs, "Ethical Sociality in Levinas and Marx"
Commentator: Andrew Tallon

2. 5:15 GHANDI-KING SOCIETY, *French Suite*

3. 5:15 SOCIETY FOR ANCIENT GREEK PHILOSOPHY, *English Suite*
Topic: Ancient Ontology
Chair: John Anton
I. Susan Sauve, "Aristotle, Teleology, and Reduction"
II. Steven Strange, "Plotinus on the Articulation of Being"

4. 5:15 SOCIETY FOR BUSINESS ETHICS, *Stuart*
Topic: Bonds and the Man's Ethical Perspectives on the Milken Case
Chair: Lisa H. Newton
I. Thomas Donaldson, "Morality and the Market for Corporate Control"
II. Thomas Dunfee, "The Law and beyond: Legal and Ethical Issues in the Enforcement of the Securities Laws"

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

III. R. Edward Freeman, "Junk: A Pragmatic View"
IV. Business Meeting (Chair: Michael Hoffman)

5. 5:15 INTERNATIONAL BERKELEY SOCIETY, *Lancaster A*
Topic: Berkeley
Chair: Phillip Cummins
I. Douglas Jesseph, "Instrumentalism and Philosophy of Mathematics in Berkeley's *Analyst*"
Commentator: David Sherry
II. Richard Van Iten, "Berkeley: Strictly Speaking"
Commentator: Daniel Flage

6. 5:15 AYN RAND SOCIETY, *Lancaster E*
Topic: Teleological Concepts
Chair: George Walsh
Harry Binswanger, "The Biological Basis of Teleological Concepts"
Commentator: William Bechtel

7. 5:15 NORTH AMERICAN KANT SOCIETY, *Austrian Suite*

8. 5:15 AMERICAN SOCIETY FOR VALUE INQUIRY, *York*
Chair: Sander Lee
I. Paul G. Kuntz, "No Rights without Duties"
II. David E. Johnson, "No Duties without Rights"
III. Thomas Magnell, "Rights and Duties"

9. 5:15 SOCIETY FOR SYSTEMATIC PHILOSOPHY, *Lancaster D*
Topic: Symposium on Dialogue and Dialectic
I. David Kolb, "What Is Open and What Is Closed in the Philosophy of Hegel"
II. John McCumber, "Dialectic and Dialogue in Hegel's Logic"

10. 5:15 PERSONALIST DISCUSSION GROUP, *Flemish Suite*
Topic: Alternative Ethical Theory
Chair: Thomas O. Buford
Speakers: Jack Padgett and Edmund L. Pincoffs

11. 5:15 SOCIETY FOR THE PHILOSOPHY OF CREATIVITY, *Spanish Suite*
Topic: Hermeneutics and Creativity
Chair: Raymond Boisvert
John Connolly, "Truth beyond Method: Gadamer on Poetry and Interpretation"
Commentator: Kathleen Wright

12. 5:15 SOCIETY FOR THE STUDY OF WOMEN PHILOSOPHERS, *Tudor*

13. 5:15 CHARLES S. PEIRCE SOCIETY, *Grecian Suite*
Chair: Hilary Putnam
I. Douglas R. Anderson, "Emersonian Elements in Peirce's Cosmology"
II. Kenneth L. Ketner, "Novel Science"
III. Peirce Society Essay Contest Award Paper

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Thursday, December 28, Early Evening, V

14. 5:15 SOCIETY FOR ASIAN AND COMPARATIVE PHILOSOPHY
Lancaster B & C

Topic: The Institution of the Family in China
Chair: David L. Hall
Speakers: Ann Birdwhistell and Yukio Kachi
Commentator: To be announced

VI. EVENING (7:30–10:30)

1. 7:30 SOCIETY FOR PHILOSOPHY AND PUBLIC AFFAIRS, *Italian Suite*
Chair: Carol Gould
Laura Purdy, "Are Pregnant Women Fetal Containers?"
Commentators: Joan Callahan and Mary Anne Warren

2. 7:30 SANTAYANA SOCIETY, *French Suite*
Chair: Paul G. Kuntz
I. Morris Grossman, "Interpreting *Interpretations*"
Commentators: Henry S. Levinson and Richard Lyon
II. Herman J. Saatkamp, Jr., "Presentation of the Critical Edition of George Santayana's *Interpretations of Poetry*"

3. 7:30 LEIBNIZ SOCIETY, *English Suite*
Chair: Margaret Wilson
I. Catherine Wilson, "Leibniz and the *Animalcula*"
II. Business Meeting

4. 7:30 SOCIETY FOR MEDIEVAL AND RENAISSANCE PHILOSOPHY
Lancaster D

Chair: Ralph McInerny
I. Bernardo Carlos Bazan, "The Turning Point of Thomas's Anthropology: *Quaestiones disputatae De anima*"
II. John F. Boler, "The 'Two Wills' Theory in Augustine, Anselm, and Scotus"

5. 7:30 SOCIETY FOR THE STUDY OF ETHICS AND ANIMALS, *Lancaster E*
Chair: Harlan B. Miller
I. Jane Duran, "Domesticated and Then Some"
Commentator: Gary E. Varner
II. Jack Weir, "Unnecessary Pain, Nutrition, and Vegetarianism"
Commentator: Steve F. Sapontzis

6. 7:30 CONFERENCE OF PHILOSOPHICAL SOCIETIES, *Cairo*
Topic: Workshop on Philosophy and Children
Chair: Telegar Satish
I. Edmund L. Pincoffs, "Philosophy for Children and Ethical Thought"
II. Martin Benjamin, "Knowledge and the Classroom"
Respondents: Matthew Lipman and Ann Margaret Sharp

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

7. 7:30 RADICAL PHILOSOPHY ASSOCIATION, *Stuart*
Topic: Feminism and Postmodernism
Chair: Diane Raymond
I. Jana Sawicki, "Foucault and Feminism: A Critical Reappraisal"
II. Douglas Kellner, "Postmodernism and Social Theory"
III. Linda Singer, "Feminism and Poststructuralism"

8. 7:30 ASSOCIATION FOR THE PHILOSOPHY OF THE UNCONSCIOUS *Spanish Suite*
Chair: Wilfried Ver Eecke
William J. Massicotte, "Why Psychoanalysis Presumes an Answer to Virtually Every Philosophical Problem"
Commentators: Edwin R. Wallace, IV and Stanley G. French

9. 7:30 PHILOSOPHERS FOR SOCIAL RESPONSIBILITY, *Flemish Suite*
Topic: Civil Wrongs
Chair: Dennis Rohatyn
Speakers: Anita Allen, Diana Axelsen, Bernard Boxill, and Burton Leiser

10. 7:30 SOCIETY FOR PHILOSOPHERS AT WORK IN THE WORLD *Grecian Suite*
Topic: Epistemological Issues in Practical Contexts
Chairs: Nance Cunningham Butler and Jean Campbell
Speakers: Elliot Cohen, Karl R. Pavlovic, and Terry Sheldahl

11. 7:30 BERTRAND RUSSELL SOCIETY, *York*
Chair: David E. Johnson
Trip McCrossin, "Russelian Objects: Unity, Complexity, and Empiricism"
Commentator: Stephen Neale

12. 7:30 WORLD PHENOMENOLOGY INSTITUTE, *Lancaster A*
Chair: Angel Medina
I. Thomas Ryba, "*The Three Movements of the Soul:* A-T. Tymieniecka's Contribution to the Philosophical Psychology of Religion"
II. Michael Saso, "The Divine Thao in Laycock's *Foundations of Theology*"
Commentator: Charles Courtney

13. 7:30 NORTH AMERICAN SOCIETY FOR SOCIAL PHILOSOPHY, *Hong Kong*
Topic: From beyond the Veil: Excursions in African American Social and Political Thought
Chair: Anita Allen
I. Robert Birt, "C.L.R. James and the Dialectics of Liberation"
II. William Lawson, "Civil Disobedience and the Social Contract"
III. Tommy Lott, "African American Social Thought in the 19th Century"

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

FRIDAY, DECEMBER 29

VII. MORNING (9:00–11:00)

1. 9:00 SOCIETY FOR MEDIEVAL AND RENAISSANCE PHILOSOPHY
English Suite

Chair: William E. Mann
I. Scott MacDonald, "The Metaphysics of Goodness and the Doctrine of the Transcendentals"
II. Mark D. Jordan, "The Transcendental Good and the Motion of Human Wills in Aquinas"

2. 9:00 SOCIETY FOR WOMEN IN PHILOSOPHY, *French Suite*
Topic: Feminist Reflections on Gender
Chair: Kate Wininger
I. Linda Nicholson, "The Category of Gender"
II. Susan Babbitt, "Myth and Realism"

3. 9:00 JASPERS SOCIETY OF NORTH AMERICA, *Italian Suite*
Chair: Harold H. Oliver
I. Kimiyo Murata-Soraci, "*Denken als Philosophie:* A Critique of Jaspers's *Nietzsche*"
Commentator: Robert Corrington
II. Business Meeting

VIII. AFTERNOON (2:15–4:15)

1. 2:15 SOCIETY FOR THE PHILOSOPHIC STUDY OF GENOCIDE AND THE HOLOCAUST, *French Suite*
Chair: Alan Rosenberg
I. Sander Lee, "The Law and Morality in War-crimes Trials"
II. Andrew Buchwalter, "The Current Dispute among German Historians over the Nazis"

2. 2:15 SOCIETY FOR THE PHILOSOPHY OF SEX AND LOVE, *Lancaster E*
Chair: Alan Soble
Timo Airaksinen, "The Style of Sade: Sex, Text, and Cruelty"
Commentator: Dorothea Olkowski

3. 2:15 AMERICAN ASSOCIATION OF PHILOSOPHY TEACHERS, *Flemish Suite*
Topic: The TV Course "Ethics in America"
Chair: Terrell Ward Bynum
Lisa H. Newton, "Socrates on the Tube: An Assessment of 'Ethics in America' as a College Ethics Course"
Commentators: Armen Marsoobian and Donald Scherer

4. 2:15 RADICAL PHILOSOPHY ASSOCIATION, *Spanish Suite*
Topic: Marxism, Rights, and Justice
Chair and Commentator: Steven Best
I. Amy Bartholomew, "Should a Marxist Believe in Marx or Rights?"
II. R. G. Peffer, "Marxism and Social Justice"

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

IX. EVENING (6:30–9:30)

1. 6:30 AMERICAN ASSOCIATION FOR THE PHILOSOPHIC STUDY OF SOCIETY
Austrian Suite
Chair: Jeremy Shearmur
I. Tara Smith, "Welfare Rights?"
Commentator: Andrew Melnyk
II. J. Roger Lee, "Rights in Context"
Commentator: John Ahrens

2. 6:30 SOCIETY FOR THE PHILOSOPHIC STUDY OF GENOCIDE AND THE HOLOCAUST, *Italian Suite*
Chair: Sander Lee
I. Laura Levitt and Robert Voyles, "Justice, Morality, and the Language of the Holocaust"
II. Business Meeting

3. 6:30 SARTRE CIRCLE, *French Suite*
Chair: Hazel E. Barnes
I. Adrian Mirvish, "Freud contra Sartre: Repression or Self-deception"
Commentator: Joseph S. Catalano
II. Kathleen V. Wider, "Sartre and Nagel on Subjectivity"
Commentator: Frithjof Bergmann

4. 6:30 APA COMMITTEE ON PRE-COLLEGE INSTRUCTION IN PHILOSOPHY
English Suite
Board Meeting

5. 6:30 AMERICAN SOCIETY FOR VALUE INQUIRY, *Tudor*
Topic: Rights and Duties
Chair: Charles Hornbeck
Christine Maitland, "Sexual Harassment in the Academic Community"

6. 6:30 HUME SOCIETY, *Lancaster E*
Chair: Eric Steinberg
Speaker: Geoffrey Sayre McCord

7. 6:30 SOCIETY FOR IBERIAN AND LATIN AMERICAN THOUGHT
Flemish Suite
Topic: Contemporary Perspectives on Latin American Philosophy
I. Roberto Vichot, "A Postmodernist Perspective on Vasconcelo's *Social Imagination*"
II. Ofelia Schutte, "The Feminist Challenge in Latin American Philosophy"
III. Business Meeting

8. 6:30 SOCIETY FOR THE STUDY OF ETHICS AND ANIMALS, *Singapore*
Chair: Harlan B. Miller
I. Dianne Romain, "Feminist Reflections in Humans and Other Domestic Animals"
Commentator: A. G. Rud

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Friday, December 29, Evening, IX

II. Wendell Stephenson, "Institutional Animal Care and Use Committees and the Moderate Position"
Commentator: Sidney Gendin

9. 6:30 ASSOCIATION FOR INFORMAL LOGIC AND CRITICAL THINKING *Cairo*
Topic: Reasoning as a Liberal Art
I. Maurice Finocchiaro, "Galileo and Scientific Reasoning"
II. Eugene Garver, "Machiavelli and the Art of Prudence"
Commentator: Alasdair MacIntyre
III. Business Meeting

10. 6:30 SOCIETY FOR SYSTEMATIC PHILOSOPHY, *Spanish Suite*
Topic: Symposium on Richard Dien Winfield's *Reason and Justice*
Speakers: Carol Gould and Robert Berman
Respondent: Richard Dien Winfield

11. 6:30 RADICAL PHILOSOPHY ASSOCIATION, *Stuart*
Topic: Roger Gottlieb's *History and Subjectivity: The Transformation of Marxist Theory*
Chair: Thomas Slaughter
Speakers: Charles W. Mills and Linda Nicholson
Respondent: Roger Gottlieb
Business Meeting

12. 6:30 NORTH AMERICAN SOCIETY FOR SOCIAL PHILOSOPHY *Grecian Suite*
Topic: The Philosophy and Theology of Liberation and Human Rights
Chair: Joseph Betz
I. Hugh Lacey, "Liberation Theology and Human Rights"
Commentator: Gene James
II. David Crocker, "Philosophy of Liberation and Human Rights"
Commentator: George Hampsch

13. 6:30 CONFERENCE OF PHILOSOPHICAL SOCIETIES, *York*
Topic: Contemporary American Philosophy: Analyst and Pluralist Critiques
Chair: John Lachs
Speakers: Nicholas Capaldi, Joseph Margolis, and Ernest Sosa

14. 6:30 INTERNATIONAL SOCIETY FOR NEOPLATONIC STUDIES, *Lancaster D*
Topic: Neoplatonism and High Scholasticism
Chair: Sheri Katz
I. Anthony J. Celano, "Albertus Magnus and Thomas Aquinas on the *De divinis nominibus*"
II. Thomas A. Losoncy, "Neoplatonic Materials in the Thought of Bonaventure"

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

III. Alan R. Perreiah, "Neoplatonic Influences on Medieval Nominalism"
IV. Richard C. Taylor, "Thomas Aquinas and the *Plotiniana Arabica*"

15. 6:30 G. E. MOORE SOCIETY, *Lancaster A*
Topic: Moore and Collingwood
Chair and Commentator: Dennis Rohatyn
Speakers: Lorraine Code, Wayne Mastin, Lionel Rubinoff, and Anthony Russell

16. 6:30 SOCIETY FOR THE STUDY OF WOMEN PHILOSOPHERS, *Manila*

17. 6:30 ASSOCIATION FOR THE PHILOSOPHY OF EDUCATION, *Hong Kong*
Topic: Do Children Think Philosophically? Thoughts after Piaget
Chair and Commentator: Matthew Lipman
Speakers: Richard Kitchener and Gareth Matthews
Business Meeting

18. 6:30 WORLD PHENOMENOLOGY INSTITUTE, *Essex A & B*
Chair: Evelyn Barker
J. N. Mohanty, "What Is Special about Phenomenology of Religion?"
Commentators: Steven Laycock and Anna-Teresa Tymieniecka

19. 6:30 SOCIETY FOR THE PHILOSOPHICAL STUDY OF MARXISM AND THE SOCIETY FOR THE ADVANCEMENT OF AMERICAN PHILOSOPHY *Lancaster B & C*
Topic: William James and Politics: Pragmatism and Marxism
Chair: John Ryder
I. Edward Hudlin, "The Impact of William James on the Thought of W.E.B. DuBois"
II. James Campbell, "William James as a Social Philosopher"
III. John Riser, "A Marxist Critique of James's Social Philosophy"

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

GROUP PARTICIPANTS

(See Appendix)

ADAMS, Robert M., University of California/Los Angeles I-3
AHRENS, John, University of Hartford IX-1
AIRAKSINEN, Timo, University of Helsinki, Finland VIII-2
ALEXANDER, Thomas, Southern Illinois University II-7
ALLEN, Anita, Georgetown University VI-13
ALLISON, David, State University of New York/Stony Brook IV-4
ALSTON, William, Syracuse University I-3
AMES, Roger T., University of Hawaii/Manoa II-2
ANDERSON, Douglas R., Wittenburg University I-2, V-13
ANTON, John, University of South Florida V-3
AXELSEN, Diana, California Lutheran University VI-9
BABBITT, Susan, Cornell University VII-2
BARKER, Evelyn, University of Maryland IX-18
BARNES, Hazel E., University of Colorado IV-5, IX-3
BARRAL, Mary Rose, Gannon College, Pennsylvania II-3
BARTHOLOMEW, Amy, Carleton University VIII-4
BAZAN, Bernardo Carlos, University of Ottawa, Canada VI-4
BECHTEL, William, Georgia State University V-6
BEINER, Ronald, University of Toronto, Canada II-4
BELL, Linda, Georgia State University III-1, IV-5
BENJAMIN, Martin, Michigan State University VI-6
BENSON, Thomas L., St. Andrews Presbyterian College, North Carolina IV-1
BERGMANN, Frithjof, University of Michigan/Ann Arbor IX-3
BERMAN, Robert, Xavier University IX-10
BERNSTEIN, Richard J., Haverford College, Pennsylvania II-7
BEST, Steven, University of Chicago VIII-4
BETZ, Joseph, Villanova University IX-12
BINSWANGER, Harry, New York, New York V-6
BIRDWHISTELL, Ann, Stockton College, New Jersey V-14
BIRT, Robert, Prairie View A & M University VI-13
BOISVERT, Raymond, Siena College, New York V-11
BOLER, John F., University of Washington VI-4
BOXILL, Bernard, University of North Carolina/Chapel Hill VI-9
BREAZEALE, Daniel, University of Kentucky IV-4
BROCKELMANS, Paul, University of New Hampshire I-5
BUCHWALTER, Andrew, University of North Florida VIII-1
BUFORD, Thomas O., Furman University V-10
BUTLER, Judith, Johns Hopkins University IV-5
BUTLER, Nance Cunningham, University of Connecticut Medical School VI-10
BYNUM, Terrell Ward, Southern Connecticut State University VIII-3
CALLAHAN, Joan, University of Kentucky VI-1
CAMPBELL, James, University of Toledo IX-19
CAMPBELL, Jean, Brooklyn, New York VI-10
CAPALDI, Nicholas, City University of New York/Queens College IX-13
CATALANO, Joseph S., Kean College of New Jersey IX-3
CELANO, Anthony J., Stonehill College, Massachusetts IX-14
CODE, Lorraine, City University of New York/Queens College IX-15
COHEN, Elliot, Indian River Community College, Florida VI-10

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

CONNELL, George, Concordia College, Minnesota II-5
CONNOLLY, John, Smith College, Massachusetts V-11
CORRINGTON, Robert, Pennsylvania State University VII-3
COURTNEY, Charles, Drew University VI-12
CROCKER, David, Colorado State University IX-12
CUMMINS, Phillip, University of Iowa V-5
DAISE, Benjamin, Hobart & William Smith Colleges, New York II-5
DANIELSON, Peter A., York University, Canada III-6
DETMER, David, Purdue University/Calumet III-1
DONALDSON, Thomas, Georgetown University V-4
DONNELLY, John, University of San Diego II-5
DOUGHERTY, Jude, Catholic University of America II-10
DOVE, Kenley R., State University of New York/Purchase II-8
DUNFEE, Thomas, Wharton School, University of Pennsylvania V-4
DURAN, Jane, University of California/Santa Barbara VI-5
DURBIN, Paul T., University of Delaware II-7
EVANS, C. Stephen, Saint Olaf College, Minnesota II-5
EZORSKY, Gertrude, City University of New York/Brooklyn College III-7
FARLEY, Edward, Vanderbilt University II-10
FINOCCHIARO, Maurice, University of Nevada IX-9
FLAGE, Daniel, University of Texas V-5
FLYNN, Thomas R., Emory University III-1
FRANKLIN, Shirley Clark, City Solicitor, Atlanta III-7
FREEMAN, R. Edward, Darden School, University of Virginia V-4
FRENCH, Stanley G., Concordia College, Minnesota VI-8
GAN, BARRY L., Saint Bonaventure University III-5
GARVER, Eugene, Saint John's University IX-9
GAVIN, William J., University of Southern Maine III-4
GENDIN, Sidney, Eastern Michigan University IX-8
GIBBS, Robert, Princeton University V-1
GORNIAK-KOCILKOWSKA, Krystyna, Temple University I-5, III-8
GOTTLIEB, Roger, Worcester Polytechnic Institute, Massachusetts IX-11
GOULD, Carol, Stevens Institute of Technology, New York VI-1, IX-10
GRANGE, Joseph, University of Southern Maine II-2
GROSSMAN, Morris, Fairfield University VI-2
HALL, David L., University of Texas/El Paso II-2, V-14
HALPER, Edward, University of Georgia II-8
HAMMONDS, Evelyn, Harvard University II-9
HAMPSCH, George, Holy Cross College, Massachusetts II-11, IX-12
HAHN, Erich, Institute of Philosophy, Academy of Sciences, German Democratic Republic II-11
HAUGLUND, John, Christopher Newport College, Virginia II-6
HAUSMAN, Carl R., Pennsylvania State University II-1
HICKMAN, Larry, Texas A & M University II-7, III-4
HOFFMAN, Michael, Bentley College, Massachusetts V-4
HOLMES, Robert L., University of Rochester III-5
HORNBECK, Charles, Keene State College, New Hampshire IV-3, IX-5
HOWARD, Michael, University of Maine/Orono II-11
HUDLIN, Edward, Babson College, Massachusetts IX-19
JAMES, Gene, Memphis State University IX-12
JESSEPH, Douglas, Franklin & Marshall College, Pennsylvania V-5

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

JOHNSON, David E., United States Naval Academy, Maryland V-8, VI-11
JORDAN, Mark D., University of Notre Dame VII-1
KACHI, Yukio, University of Utah V-14
KATZ, Sheri, Spring Hill College, Alabama IX-14
KATZNER, James S., Wheeler School II-6
KATZNER, Louis, Bowling Green State University II-6
KELLNER, Douglas, University of Texas/Austin VI-7
KETNER, Kenneth L., Texas Tech University V-13
KITCHENER, Richard, Colorado State University IX-17
KLUBACK, William, City University of New York/Kingsborough Community College III-8
KOLB, David, Bates College, Maine V-9
KRELL, David Farrell, DePaul University IV-4
KRONEGGER, Marlies, Michigan State University II-10
KUNTZ, Paul G., Emory University V-8, VI-2
LACEY, Hugh, Swarthmore College, Pennsylvania IX-12
LACHS, John, Vanderbilt University IX-13
LANGLEY, Raymond L., Manhattanville College, New York I-2
LAW, Walter, Emory University I-5
LAWSON, William, University of Delaware VI-13
LAYCOCK, Steven, University of Hawaii/Manoa IX-18
LEE, J. Roger, California State University IX-1
LEE, Kwang-Sae, Kent State University IV-1
LEE, Sander, Keene State College, New Hampshire III-3, IV-3, V-8, VIII-1, IX-2
LEISER, Burton, Pace University VI-9
LEVINSON, Henry S., University of North Carolina/Greensboro VI-2
LEVITT, Laura, Emory University IX-2
LINDBECK, George, Yale University I-3
LINGIS, Alphonso, Pennsylvania State University IV-4
LIPMAN, Matthew, Montclair State College, New Jersey II-6, VI-6, IX-17
LIST, Charles, State University of New York/Plattsburgh IV-3
LONGINO, Helen, Mills College, California II-9
LOSONCY, Thomas A., Villanova University IX-14
LOTT, Tommy, University of Massachusetts/Boston VI-13
LOUGHNEY, John A., Westfield State College, Massachusetts IV-3
LUBAN, David, University of Maryland IV-2
LYON, Richard, Hampshire College, Massachusetts VI-2
MACDONALD, Scott, University of Iowa VII-1
MACINTYRE, Alasdair, University of Notre Dame IX-9
MACHAN, Tibor R., Auburn University III-6
MAGID, Carolyn, Bentley College, Massachusetts III-7
MAGNELL, Thomas, Smith College, Massachusetts III-3, V-8
MAITLAND, Christine, National Education Association IX-5
MANN, William E., University of Vermont VII-1
MARGOLIS, Joseph, Temple University IX-13
MARSOOBIAN, Armen, Southern Connecticut State University VIII-3
MASSICOTTE, William J., McGill University, Canada VI-8
MASTIN, Wayne, United States Military Academy, New York IX-15
MATTHEWS, Gareth, University of Massachusetts/Amherst II-6, IX-17
McCORD, Geoffrey Sayre, University of North Carolina/Chapel Hill IX-6
McCROSSIN, Trip, Stanford University VI-11

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

McCUMBER, John, Northeastern University V-9
McINERNY, Ralph, University of Notre Dame VI-4
McKENZIE, Nancy, Vassar College, New York VI-9
MEDINA, Angel, Georgia State University VI-12
MELNYK, Andrew, Corpus Christi College, Texas IX-1
MILLER, Harlan B., Virginia Polytechnic Institute & State University VI-5, IX-8
MILLS, Charles W., University of Oklahoma IX-11
MIRVISH, Adrian, California State University/Chico IX-3
MOHANTY, J. N., Temple University IX-18
MORKOVSKY, Mary Christine, Wadhams Hall Seminary-College, New York V-1
MORRIS, Phyllis S., Oberlin College, Ohio III-1
MURATA-SORACI, Kimiyo, Boston University VII-3
MURPHY, Jeffrie, Arizona State University III-5
MURPHY, Julien S., University of Southern Maine III-1
NARVESON, Jan, University of Waterloo, Canada III-6
NEALE, Stephen, Princeton University VI-11
NEWTON, Lisa H., Fairfield University V-4, VIII-3
NICHOLSON, Linda, State University of New York/Albany VII-2, IX-11
OLIVER, Harold H., Boston University III-8, VII-3
OLKOWSKI, Dorothea, College of Charleston, South Carolina VIII-2
OLSON, Alan, Boston University I-5, III-8
PADGETT, Jack, Albion College, Michigan V-10
PAVLOVIC, Karl R., Gaithersburg, Maryland VI-10
PEFFER, R. G., University of San Diego VIII-4
PERREIAH, Alan R., University of Kentucky IX-14
PINCOFFS, Edmund L., University of Texas/Austin V-10, VI-6
POJMAN, Louis, University of Mississippi II-5
POTTER, Elizabeth, Hamilton College, New York II-9
PURDY, Laura, Hamilton College, New York VI-1
PUTNAM, Hilary, Harvard University III-4, V-13
PUTNAM, Ruth Anna, Wellesley College, Massachusetts III-4
QUINN, PHILIP, University of Notre Dame I-3
RAYMOND, Diane, Simmons College, New York VI-7
RISER, John, University of Central Florida IX-19
ROHATYN, Dennis, University of San Diego VI-9, IX-15
ROMAIN, Dianne, Sonoma State University IX-8
ROSE, MARY CARMEN, Goucher College, Maryland II-1
ROSENBERG, ALAN, City University of New York/Queens College VIII-1
ROWNTREE, Stephen C., Loyola College, Louisiana II-5
RUBINOFF, Lionel, Trent University IX-15
RUD, A. G., Western Carolina University IX-8
RUSSELL, Anthony, Loras College, Iowa IX-15
RYBA, Thomas, Michigan State University VI-12
RYDER, John, State University of New York/Cortland IX-19
SAATKAMP, Jr., Herman J., Texas A & M University VI-2
SAPONTZIS, Steve F., California State University/Hayward VI-5
SASO, Michael, University of Hawaii/Manoa VI-12
SATISH, Telegar, Coppin State College, Maryland VI-6
SAUVE, Susan, Harvard University V-3
SAWICKI, Jana, University of Maine II-9, IV-5, VI-7
SCHERER, Donald, Bowling Green State University VIII-3

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

SCHUTTE, Ofelia, University of Florida II-11, IX-7
SHAPIRO, Gary, University of Kansas IV-4
SHARP, Ann Margaret, Montclair State College, New Jersey VI-6
SHEARMUR, Jeremy, George Mason University III-6, IX-1
SHELDAHL, Terry, Savannah State College, Georgia VI-10
SHER, George, University of Vermont I-4
SHERRY, David, Flagstaff, Arizona V-5
SHUE, Henry, Berkshire, New York VI-9
SIEGEL, Harvey, University of Miami II-6
SINGER, Linda, Miami University VI-7
SLAUGHTER, Thomas, Bentley College, Massachusetts IX-11
SLEEPER, Ralph, City University of New York/Queens College III-4
SMITH, Joel R., Skidmore College, New York IV-1
SMITH, Tara, University of Texas/Austin IX-1
SOBLE, Alan, University of New Orleans VIII-2
SOSA, Ernest, Brown University IX-13
SPITZER, Robert, Georgetown University II-3
STEINBERG, Eric, City University of New York/Brooklyn College IX-6
STEPHENSON, Wendell, Lewis & Clark College, Oregon IX-8
STERBA, James, University of Notre Dame III-5
STRANGE, Steven, University of Pittsburgh V-3
SULLIVAN, Patrick, Charlottesville, Virginia I-2
TALLON, Andrew, Marquette University V-1
TAYLOR, Richard C., Marquette University IX-14
THAYER, H. S., City University of New York/City College III-4
TYMIENIECKA, Anna-Teresa, Belmont, Massachusetts IX-18
VAN ITEN, Richard, Iowa State University V-5
VARNER, Gary E., Washington University VI-5
VER EECKE, Wilfried, Georgetown University VI-8
VICHOT, Roberto, Texas A & M University IX-7
VOYLES, Robert, Emory University IX-2
WALLACE, IV, Edwin R., Medical College, Georgia VI-8
WALSH, George, Salisbury State University V-6
WALSH, Margaret, Bradford College, Massachusetts IV-5
WALTERS, Greg, Saint Mary's University III-8
WARD, David, Widener University IV-3
WARNER, Stuart, Roosevelt University III-6
WARREN, Mary Anne, San Francisco State University VI-1
WEINSTEIN, Mark, Montclair State College, New Jersey II-6
WEIR, Jack, Hardin-Simmons University VI-5
WENISCH, Fritz, University of Rhode Island/Kingston II-1
WIDER, Kathleen V., University of Michigan/Dearborn IX-3
WILLIAMS, Jay G., Hamilton College, New York IV-1
WILLIAMS, Robert R., Hiram College, Ohio II-10
WILSON, Catherine, University of Oregon VI-3
WILSON, Margaret, Princeton University VI-3
WINFIELD, Richard Dien, University of Georgia IX-10
WININGER, Kate, University of Southern Maine VII-2
WRIGHT, Kathleen, Haverford College, Pennsylvania V-11
YOUNG, Iris, Worcester Polytechnic Institute, Massachusetts III-7
ZIOLKOWSKI, Eric J., Lafayette College, Pennsylvania II-5

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVI, NO. 11, NOVEMBER 1989

FOUCAULT AND THE PARADOX OF BODILY INSCRIPTIONS*

THE position that the body is constructed is one that is surely, if not immediately, associated with Michel Foucault. The body is a site where regimes of discourse and power inscribe themselves, a nodal point or nexus for relations of juridical and productive power. And, yet, to speak in this way invariably suggests that there is a body that is in some sense there, pregiven, existentially available to become the site of its own ostensible construction. What is it that circumscribes this site called "the body"? How is this delimitation made, and who makes it? Which body qualifies as "the" body? What establishes the 'the', the existential status of this body? Does the existent body in its anonymous universality have a gender, an unspoken one? What shape does this body have, and how is it to be known? Where did "the body" come from? To claim that "the body is culturally constructed" is, on the one hand, to assert that whatever meanings or attributes the body acquires are in fact culturally constituted and variable. But note that the very construction of the sentence confounds the meaning of 'construction' itself. Is "the body" ontologically distinct from the process of construction it undergoes? If that is the case, then it would appear that "the body," which is the object or surface on which construction occurs, is itself prior to construction. In other words, "the body" would not be constructed, strictly considered, but would be the occasion, the site, or the condition of a process of construction only externally related to the body that is its object. In effect, the statement, 'the body is constructed', refuses to allow that the indefinite article is itself a construction that calls for a genealogical account. But perhaps the

* To be presented in an APA symposium on the Thought of Michel Foucault, December 30. Ladelle McWhorter and David F. Gruber will be co-symposiasts; see this JOURNAL, this issue, 608–614 and 615–621, respectively, for their contributions.

0022-362X/89/8611/601–607 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

referential claim of the statement is rhetorical, that is, treating the body named within the sentence not as an object that receives a construction, i.e., as a substantive thing that might take on a contingent attribution. Perhaps instead it is necessary to read the statement in a self-referential way, that is, as asserting that any reference to "the" body in its indefiniteness is of necessity a construction, one that is open to a genealogical critique.

Within a number of texts, Foucault clearly questions whether there is a "materiality" to bodies which is in any sense separable from the ideational or cultural meanings that constitute bodies within specific social fields. In *The History of Sexuality: Vol. I*,[1] he claims that the body is a site of culturally contested meanings, and that "sex," what we might be tempted to rank among the most factic aspects of bodily life, is itself an "imaginary point," the consequence of a materiality fully "invested" with ideas. In "Nietzsche, Genealogy, History,"[2] Foucault clearly writes, ". . . nothing in man (sic) —not even his body—is sufficiently stable to serve as the basis for self-recognition or for understanding other men (sic)" (*ibid.*, p. 153). Foucault's efforts to describe the mechanism by which bodies are constituted as cultural constructions, however, raises the question of whether there is in fact a body which is external to its construction, invariant in some of its structures, and which, in fact, represents a dynamic locus of resistance to culture per se.

This last claim is no doubt surprising for anyone who has read Foucault's critique of the twin evils of liberationist sexuality and psychoanalytic theory in *The History of Sexuality: Vol. I*. In that text Foucault claimed quite clearly that there could be no body before the law, no sexuality freed from relations of power. If the body in its indefinite generality, however, proves to be a point of dynamic resistance to culture per se, then this body is not culturally constructed, but is, in fact, the inevitable limit and failure of cultural construction. I shall argue in the following that, whereas Foucault wants to argue—and does claim—that bodies are constituted within the specific nexus of culture or discourse/power regimes, and that there is no materality or ontological independence of the body outside of any one of those specific regimes, his theory nevertheless relies on a notion of genealogy, appropriated from Nietzsche, which conceives the body as a surface and a set of subterranean "forces" that are, indeed, repressed and transmuted by a mechanism of cultural construction external to that body. Further, this mechanism of

[1] Robert Hurley, trans. (New York: Vintage, 1980).
[2] In *Language, Counter-Memory, Practice: Selected Essays and Interviews by Michel Foucault*, Donald F. Bouchard, trans., Sherry Simon and Donald F. Bouchard, eds. (Ithaca: Cornell, 1977).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

cultural construction is understood as "history," and the specific operation of "history" is understood, and understood problematically, as inscription. Indeed, I shall try to show that, for Foucault, not unlike for Kafka in *The Penal Colony,* the cultural construction of the body is effected through the figuration of "history" as a writing instrument that produces cultural significations—language—through the disfiguration and distortion of the body, where the body is figured as a ready surface or blank page available for inscription, awaiting the "imprint" (*ibid.,* 148) of history itself. Although Foucault appears to argue that the body does not exist outside the terms of its cultural inscription, it seems that the very mechanism of "inscription" implies a power that is necessarily external to the body itself. The critical question that emerges from these considerations is whether the understanding of the process of cultural construction on the model of "inscription"—a logocentric move if ever there was one—entails that the "constructed" or "inscribed" body have an ontological status apart from that inscription, precisely the claim that Foucault wants to refute.

Within "Nietzsche, Genealogy, History," Foucault describes the body through a series of metaphors and figures, yet predominantly as a "surface," a set of multidirectional "forces," and as the scene or site of a cultural inscription. He writes: "the body is the inscribed surface of events" (*ibid.*). The task of genealogy, he claims, is "to expose a body totally imprinted by history" (*ibid.*). His sentence continues, however, by referring to the goal of "history" as the "destruction of the body" (*ibid.*). This remark paradoxically recalls Freud's description of "civilization," that juridical structure of power exercised through the repression and sublimation of primary drives that Foucault seeks to criticize in *The History of Sexuality: Vol. I.* Forces and impulses with multiple directionalities are precisely what "history" both destroys and preserves through the *Entstehung* (historical emergence) of inscription. As "a volume in perpetual disintegration," Foucault writes, the body is always under seige, suffering destruction by the very terms of history. History is understood as that creator of values and meanings, that signifying practice, which requires the subjection of the body in order to produce the speaking subject and its significations. Described through the language of surface and force, this is a body weakened through what Foucault in a radically ahistorical way terms the "single drama" of history, a drama or a structure of repressive signification which requires a subjugation and inscription of the body for the creation of new values (*ibid.,* p. 150). This is not the modus vivendi of one kind of history rather than another, but rather "history" (*ibid.,* p. 148) in its essential repressive/generative gesture.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Although Foucault writes that the body is not stable and cannot serve as a common identity among individuals cross-culturally or transhistorically, he nevertheless points to the constancy of cultural inscription as a "single drama," suggesting that this drama of historical "inscription" enjoys the very universality denied to the body per se. If the creation of values, that signifying practice of history, requires the destruction of the body, much as the instrument of torture in Kafka's *Penal Colony* destroys the body on which it writes, then there must be a body prior to that inscription, stable and self-identical, subject to and capable of that sacrificial destruction. In a sense, for Foucault, as for Nietzsche, cultural values emerge as the result of an inscription *on* the body, where the body is understood as a medium, indeed, a blank page, an unusual one, to be sure, for it appears to bleed and suffer under the pressure of a writing instrument. In order for this inscription to signify, however, the body-as-medium must itself be destroyed and transvaluated into a sublimated (to use Freud's language) or "transvaluated" (to use Nietzsche's language) domain of values. Implicit to this description of the creation of cultural values is the notion of history as a relentless writing instrument, and the body as the medium that must be destroyed and transfigured in order for "culture" to emerge.

By maintaining a body prior to its cultural inscription, Foucault appears to assume a materiality to the body prior to its signification and form. In separate contexts, Foucault appears paradoxically to criticize both Freud and Nietzsche for assuming a prediscursive ontology of the body and its drives. In *The History of Sexuality: Vol. I,* he argues that desires that psychoanalysis claims are repressed are, in fact, produced by psychoanalysis as a juridical practice. Indeed, "repressed desire," he appears to suggest, is a consequence of psychoanalytic discourse, a delectable discursive production, a contemporary fabrication and organization of sexuality. Quite directly, in fact, he argues against the conceptualization of a juridical and repressive law that presumes that there is a "rebellious energy" (p. 81) that temporally and ontologically precedes repression. The effect of his genealogical critique of the repressive hypothesis is to conceive the discursive figuration of drives and/or desires as simultaneous with or consequent to the law, rather than its antecedent presupposition. Interestingly, Foucault provides, in *Discipline and Punish: The Birth of the Prison,*[3] an implicit yet parallel critique of Nietzsche's presumption of a prediscursive instinctuality. In *The Genealogy of Morals* (section II), Nietzsche argues that the multi-

[3] Alan Sheridan, trans. (New York: Vintage, 1979).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

plicitous and life-affirming instincts are negated and internalized through the prohibitive effects of slave-morality. Foucault, however, challenges the language of internalization and the distinction between inner psychic and external social space which that language implies.

Foucault objected to what he understood to be the psychoanalytic belief in the "inner" truth of sex in *The History of Sexuality: Vol. I.* In *Discipline and Punish,* he similarly refuses the doctrine of internalization in his account of the subjection and subjectivation of criminals. The prohibitive law is not taken into the body, internalized or incorporated, but rather is written *on* the body, the structuring principle of its very shape, style, and exterior signification. Consider the generative play of surface significations and the refusal to engage a concept of interior psychic space in the following quotation: "the body is . . . directly involved in a political field; power-relations have an immediate hold on it; they invest it, mark it, train it, torture it, force it to carry out tasks, to perform ceremonies, to emit signs" (*ibid.,* p. 25).

In a sense, *Discipline and Punish* can be read as Foucault's effort to reconceive Nietzsche's doctrine of internalization as a language of inscription. In the context of prisoners, Foucault writes, the strategy has not been to enforce a repression of their criminal impulse, but to compel their bodies to signify the prohibitive law as their manifest essence, style, and necessity. That law is not literally internalized, but incorporated *on* bodies; there the law is manifest as a sign of the essence of their selves, the meaning of their soul, their conscience, the law of their desire. In effect, the law is fully dissimulated into the body as such; it is the principle that confers intelligibility on that body, the sign by which it is socially known. The juridical law no longer appears external to the bodies it subjects and subjectivates. "It would be wrong," Foucault writes, "to say that the soul is an illusion, or an ideological effect. On the contrary, it exists, it has a reality, it is produced permanently *around, on, within,* the body by the functioning of a power that is exercised on those that are punished . . ." (*ibid.,* p. 29; my italics). The figure of the interior soul understood as "within" the body is produced through its inscription *on* the body; indeed, the soul is inscribed on the surface, a signification that produces on the flesh the illusion of an ineffable depth. The soul as a structuring invisibility is produced in and by signs that are visible and the corporeal. Indeed, the soul requires the body for its signification, and requires also that the body signify its own limit and depth through corporeal means. Furthermore, the body must signify in a way that conceals the very fact of that signifying, indeed that

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

makes that signifying practice appear only as its reified "effect," that is, as the ontological necessity of a defining and immaterial internality and depth.

The effect of the soul as a structuring inner space is produced through the signification of a body as a vital and sacred enclosure. The soul is precisely what the body lacks; hence, the body presents itself as a signifying lack. That lack which *is* the body signifies the soul as that which cannot show. In this sense, then, the soul is a surface signification that contests and displaces the inner/outer distinction itself, a figure of interior psychic space inscribed *on* the body as a social signification that perpetually conceals itself as such. In Foucault's terms, the soul is not imprisoned by or within the body but, inversely, "the soul is the prison of the body" (*ibid.*, p. 30).

The critical power of Foucault's analysis assumes that only under certain conditions of power and discourse do bodies get signified and regulated in the ways that he describes in *Discipline and Punish* and in *The History of Sexuality: Vol. I.* In "Nietzsche, Genealogy, and History," however, Foucault "confesses" his metaphysical commitments in such a way that the critical power of his genealogical critique is severely undermined. In that essay, he makes clear that "subjection," though historically specific in its modalities, is also the essential and transhistorical precondition of "history" writ large; indeed, he makes clear that this significatory or generative subjection is the essential gesture of a singular history, its one infinitely repeatable "drama."

In the introduction to *The Order of Things,*[4] Foucault rather defensively suggests that, although structuralism exists as the reigning philosophical ethos of his time, and his own work invariably engages the philosophical vocabulary of structuralism, he nevertheless does not feel constrained by its terms. Although I would not want to reduce Foucault's work to a structuralist position, I would argue that the notion of signification as a universal cultural law that produces meanings through the subjection and disfiguration of the materiality of bodies seems fully derivable from Claude Lévi-Strauss's analysis of the structure of kinship exchange. For Lévi-Strauss, the incest taboo institutes identities along the axis of sexual difference and, hence, represses bodies in order to produce highly structured kinship relations. The materiality of "bodies" for Lévi-Strauss constitutes "nature" or "the raw," whereas the prohibitions that repress sexuality and create social organization are "culture" and the "cooked." Whereas Foucault appears to criticize precisely those binary oppositions encoded by structuralism as the universal

[4] New York: Random, 1971.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tensions of anthropology, he appears to reengage those oppositions in his own descriptions of how historical meanings come into being. That history is "inscribed" or "imprinted" onto a body that is not history suggests not only that the body constitutes the material surface preconditional to history, but that the deregulation and subversion of given regimes of power are effected by the body's resistance against the workings of history itself. In other words, Foucault appears to have identified in a prediscursive and prehistorical "body" a source of resistance to history and to culture, where history and culture are finally and paradoxically conceived in juridical terms. That this is contrary to Foucault's stated program to formulate power in its generative as well as juridical modes seems clear. Yet his statements on "history" appear to undermine precisely the insight into the constructed status of the body which his studies on sexuality and criminality were supposed to establish.

Because the distinction between the historical act of inscription and the body as surface and resistance is presupposed in the task of genealogy as he defines it, the distinction itself is precluded as an object of genealogical investigation. Indeed, the distinction not only operates as an uncritically accepted and implicitly formulated premise of his argument, but it ends up undermining the central point that his argument concerning the constructed status of bodies was supposed to prove. Occasionally, in his *The History of Sexuality: Vol. I,* or in his brief introduction to the journals of Herculine Barbin, the nineteenth-century hermaphrodite, Foucault seeks recourse to a prediscursive multiplicity of bodily forces that break through the surface of the body to disrupt the regulating practices of cultural coherence imposed upon that body by a regulatory regime, understood as some vicissitude of "history." If the presumption of a source of potential precategorial disruption is refused, and if the very notion of the body as surface externally related to the act of inscription is subjected to a genealogical critique, would it be possible to give a Foucaultian account of the demarcation of bodies as such as a signifying practice? How would "the" body as cultural or discursive practice be described?

What is clear is that inscription would be neither an act initiated by a reified history nor the performative accomplishment of a master historian who produces history as he writes it. The culturally constructed body would be the result of a diffuse and active structuring of the social field with no magical or ontotheological origins, structuralist distinctions, or fictions of bodies, subversive or otherwise, ontologically intact before the law.

JUDITH BUTLER
Johns Hopkins University

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

CULTURE OR NATURE? THE FUNCTION OF THE TERM 'BODY' IN THE WORK OF MICHEL FOUCAULT*

MICHEL FOUCAULT is often taken, mistakenly, to be a partisan of some form of human liberation. To some extent, the mistake is understandable; his books, essays, and interviews constitute challenges to the forces and authorities that constrain us, and they frequently call for resistance to the regimes of power in which we find ourselves. But the problem with reading Foucault as a liberationist is that, after he places in question sexuality, sanity, health, and selfhood, he leaves us with nothing to liberate. Nevertheless, those who cannot understand resistance to authoritative power as anything other than a drive for liberation persevere in reading Foucault as a liberationist and simply posit something (they claim to have caught sneaking in Foucault's backdoor) that can serve as the anchor and justification for his supposedly liberationist moves. Sometimes what such readers claim to catch sneaking in is the body, or more precisely, a notion of a transhistorical, natural body that has been and is currently repressed.

It is possible to pick out passages that would support such a reading. At the end of *The History of Sexuality: Vol. I,* Foucault[1] writes: "The rallying point for the counterattack against the deployment of sexuality ought not to be sex-desire, but bodies and pleasures" (*ibid.*, p. 157). We might take this passage as evidence that in fact Foucault is calling for a liberation of the body per se as opposed to the sexualized body. There are numerous passages in *Discipline and Punish: The Birth of the Prison*[2] which can be similarly construed. For example, he writes that eighteenth-century Europe's disciplinary regimes did not constitute "the first time that the body had become the object of such imperious and pressing investments; in every society, the body was in the grip of very strict powers, which imposed on it constraints, prohibitions or obligations" (*ibid.*, p. 136). In the eighteenth century, however, what was being formed "was a policy of coercions that act upon the body, a calculated manipulation of its elements, its gestures, its behaviour. The human body was entering a machinery of power that explores, breaks it down, and rearranges it" (*ibid.*, p. 138). From these passages, it would seem that Foucault is

* To be presented at an APA symposium on the Thought of Michel Foucault, December 30. Judith Butler and David F. Gruber will be co-symposiasts; see this JOURNAL, this issue, 601–607 and 615–621, respectively, for their contributions.

[1] Robert Hurley, trans. (New York: Random House, 1978).

[2] Alan Sheridan, trans. (New York: Random House, 1975).

0022-362X/89/8611/608–614
© 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

calling on us to free the body, to release it from the matrices of machine-like forces that hold it in their grasp.

There are many passages, however, that tell against this reading of Foucault. In *Discipline and Punish,* Foucault implies that "the natural body," just as the mechanized and docile body, is a product of discipline: "Disciplinary power has as its correlative an individuality that is not only analytical and 'cellular', but also natural and 'organic' " (*ibid.*, p. 156). In his beautiful essay, "Nietzsche, Genealogy, History,"[3] Foucault's many remarks on the body, taken together, indicate that the body is to be thought, not as a stable, unitary volume or a constant set of rhythmic processes unaffected by historical change, but rather as a locus or point of intersection of historical matrices of power: "The body is the inscribed surface of events (traced by language and dissolved by ideas), the locus of a dissociated Self (adopting the illusion of a substantial unity), and a volume in perpetual disintegration" (*ibid.*, p. 148). And again: "nothing in man—not even his body—is sufficiently stable to serve as the basis for self-recognition or for understanding other men" (*ibid.*, p. 153).

What, then, is this body to which Foucault refers? It seems to have none of the characteristics required for us to form a conception of it; it is neither singular (rather "it" "is" a plurality of perhaps quite dissimilar events), nor is it stable (so certain knowledge of it is impossible). It would seem, then, that the term 'body' in Focault's discourse is not a name for any *thing* at all. Perhaps we ought to reformulate our question: How is the term 'body' functioning in Foucault's discursive practice?

In order to understand what is occurring through Foucault's use of this term, perhaps we must first understand what Foucault is not doing with it. As stated above, Foucault is not claiming that body, a natural object, stands somehow in opposition to that which culture would impose upon it. In Foucault's discourse, the term 'body', then, is not to be situated within the dichotomy nature/culture. Nevertheless, given the nature/nurture debates raging in our day, our tendency is to hear Foucault's work through our preoccupation with that issue. Therefore, before discussing Foucault's work any further, let us examine—and thus perhaps eliminate—this obstacle to hearing Foucault on his own terms.

The dichotomy nature/culture is extraordinarily powerful. Through much of human history, if we are to believe our colleagues

[3] In *Language, Counter-Memory, Practice: Selected Essays and Interviews by Michel Foucault,* Donald F. Bouchard, trans., Sherry Simon and Donald F. Bouchard, eds. (Ithaca: Cornell, 1977).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

in anthropology, the opposition between nature and culture has ordered our thinking and dictated the structures of our lives. Nature we take as the given; culture as the humanly contrived. That which participates in nature exclusively is not in human control and as such is untrustworthy, in need of placation, evocative of fear. That which participates in culture escapes from the whims of nature to some, perhaps high, degree. Culture is our refuge, our protection; it is what we can claim and use and understand as truly our own.

Furthermore, anthropologists report, culture is generally assigned a high value that nature is almost universally denied. In anthropological literature, statements such as the following are commonly made:

> The [widespread ritualistic] celebration of motherhood and female sexuality implies a definition of womankind in terms of nature and biology; it traps women in their physical being, and thereby in the very general logic which declares them less capable of transcendence and of cultural achievement than men.[4]

That nature and culture are separate realms is very evident in such passages, as is the common understanding of culture as the more valuable of the two. The passage implies that whatever is associated primarily with nature is far less important than anything designated as a cultural achievement, and, if women are identified with nature as opposed to culture, then women are denigrated.

Some scholars point out that the strong tendency among various cultures to link human biology—particularly female functions such as menstruation, pregnancy, and childbirth, but also common processes such as aging and death—with a feared, hated, and devalued natural realm devalues women unfairly and results in masculinistic or male supremacist societies. These thinkers call into question the work of anthropologists such as Weston La Barre,[5] who happily asserts that women are not cultural animals at all, as are men, but are biological beings entirely. La Barre claims that culture, as opposed to nature, is totally and unshakeably male.

> It may be true that some of the arts—probably pottery, agriculture, and weaving—are the inventions of women. But principle or generalization

[4] Michelle Zimbalist Rosaldo and Jane Monnig Atkinson, "Man the Hunter and Woman: Metaphors for the Sexes in Ilongot Magical Spells," in *Reader in Comparative Religion: An Anthropological Approach*, 4th ed., William A. Lessa and Evon Z. Vogt, eds. (New York: Harper & Row, 1979), p. 139. This article was reprinted from *The Interpretation of Symbolism*, Roy Willis, ed. (London: Dent & Sons, 1975), pp. 43–75.

[5] *The Human Animal* (Chicago: University Press, 1954).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

is a male artifact: the *logos* that is the endless preoccupation of male metaphysics. What connects father and son, male and male, is the mystery of *logos* and logos alone: logos as the literal "word" which conveys linguistic meaning and understanding; logos as laws, agreements, rules, and regularities of behavior; logos as the implicit means and substance of common understanding and communication, and of cultural joining in the same styles of thinking; and logos as shared pattern, within which father can identify with son and permit his infancy, within which son can identify with father and become a man, and within which a male can perceive and forgive the equal manhood of his fellow-man (*ibid.*, p. 212).

Clearly for anthropologists like La Barre, culture is what is to be celebrated, for culture, and culture alone, gives us our humanity—the ability to communicate, to form communities wherein individuals are treated with justice and respect, to form families and traditions, and to produce works of art. And culture clearly stands over against nature and defies it, outsmarts it by resisting its power to dissolve and disperse in the oblivion of animalistic forgetfulness and death.

Of course, as feminists point out, if we assume that the oppositional strings are as follows—nature/body/female versus culture/biological transcendence/male—then there is a certain difficulty to be overcome, namely, that men are bodily, too. Men also participate in reproduction; men also die. Perhaps men ought not devalue nature so.

But La Barre suggests a different alternative—that men resolve the issue of their bodily being culturally; they "enculturate" their flesh with ritual and adornment, so that we must understand male bodies as thoroughly culturally inscribed.

> Primitive men know rightly that women can make children with their bodies. But it takes men to make men, that is, members of the tribe. Hence that universal preoccupation among men everywhere with initiation, that mysterious male re-birth of the youth into full membership in the society of men (*ibid.*).

Thus, the male body becomes a cultural artifact, while the female body remains rooted in and bound by its natural functioning.

Many feminists, and others, suggest that this solution ultimately cannot succeed. Denial of one's fleshly finitude results, not in transcendence toward truth and eternality, but rather in self-hatred, violence, and destruction of our habitat, the natural world. Instead of seeking salvation in cultural mechanisms of escape from pure bodily existence, we need to learn to value nature and to respect it.

Those who put forth such criticisms surely are right that the op-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

position nature/culture has tended toward hierarchy with culture placed on top and that this hierarchy of value has resulted, not only in our society's lack of respect for femaleness and its dread of and preoccupation with dying and death, but also its reckless manipulation and destruction of the natural world. There is much to be said in favor of this kind of claim and its attendant fields of research. Some —perhaps many—of those working in these fields frequently replicate, however, the hierarchical dualism they so vehemently disclaim. Too often the hierarchical valuation that places culture over nature is simply reversed so that nature becomes the more valued term, but the forces that put nature and culture in opposition to begin with are left firmly in their place.

If Foucault's calls for resistance could be best understood as calls to liberate some sort of prediscursive or natural body, we could situate his discourse among those which reverse the culture/nature hierarchy rather than among those which challenge that dualism at its root. But the term 'body' in Foucault's discourse is not functioning as the name of something naturally given, of a pure truth to be liberated; for, as Foucault reminds us, the natural body is itself a discursive phenomenon.

Many readers of Foucault are deaf to the significance of this point. They simply do not hear Foucault when he says that nature itself is historically and discursively produced. Those of us who feel some sympathy with one or another "back to nature" movement have a tendency to see, or at least to sense, a bodily existence untouched by the vicissitudes of a politically organized world and lurking behind or beneath the disciplinary regimes that modern forms of power impose. We dream—and think we know—of a body that is apolitical, ahistorical, a body that is clean. We tend to believe that there exists somewhere, just under the surface of the heavy make-up or the studied social grace, or buried deep beneath the ascetic practices of our appearance-conscious age, a natural organism—yes, more or less repressed, more or less distorted, but still the bearer of the standard of healthy functioning, the physiologically right. We tend to situate the body as the true, genuine, and opposite term in relation to the illusion of culturally produced bodily appearances so in evidence among us now—in opposition to the steroid-produced football player, the city dweller with the salon-created tan, the anorexic's Giacometti frame. And we dream of re-inhabiting that body, of rehabilitating it, of polishing its tarnished exterior away and letting it shine forth again. We dream of a liberation of pure and everlasting flesh.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

But let us pay good Foucaultian attention to this dream. What speaks in it? What power is it articulating? And what does it dispossess? Is this dream so different from La Barre's dream of transcending the natural realm? La Barre's enculturated man was a man lifted out of the realm of the animal, of the transitory, of that which will not last. This dream of a return to nature, of which the body for us is a part, is, I submit, a continuation of the very same dream. After all, nature is not just that which culture denigrates. Nature is also that from which man falls. *Nature* conjures up images of a forsaken, half-forgotten paradise, an order more perfect than anything a human can create, a place where male and female lived without a care. The flip side of the hierarchical order that places culture over nature as her tamer and border patrol is the hierarchical order—so compelling after industrialization and world wars have brought the disappearance or destruction of culture after culture after culture —of *nature* over culture, of *nature* as that which can be counted upon, as that which, unchanging, endures. A return to nature in our day and age is just another gesture in the general movement away from that which *dies*. Nature is now the eternal; culture, that which always passes away.

Foucault, in his use of the word 'body', refuses that dream, and, if we are to understand his work, we must refuse it as well. We must resist our tendency to align body with some broad concept of "the natural" as opposed to "the humanly produced." In Foucault's work, the term 'body' speaks resistance to any attempt at identification, to any attempt to capture and contain. Body is precisely that which changes, grows, degenerates, dies, decays, that which is never stable, never fully predictable, that which is opaque, elusive, and unknown. 'Body' in Foucault's discourse does not function as a name of some thing, as a linguistic symbol representing a real object out there somewhere obediently sitting still while it is referred to. Body is what will not rest in the self-same, or, perhaps better stated, 'body' names the movements of differing which repeatedly disrupt discourses of identity and strike fear in those forces whose source is our belief in the value of the timeless and the reality of some eternal presence of the same. Thus does body in this discourse of Foucault's stand in opposition, not to culture, but to the discourse that gives us nature/culture dualisms and the disputes to which they give rise. And thus does resistance in Foucault's discourse become something altogether different from a liberation of repressed true being.

Foucault is not a liberationist of any sort. He is not simply trying, like so many others, to revalue the devalued half of any oppositional

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

 Digitized by Arya Samaj Foundation Chennai and eGangotri

couplet. He is not trying to say, "Down with ordered society! Up with the animal in us!" For the animal, too, is an historically situated event. Nature, too, has a history—histories—that can be genealogically traced. In Foucault's work, 'body' has no trans- or extra-historical meaning; it has, simply, a function. It functions as a scattering point, a locus of dispersion for all that proponents of culture as the escape from death or of nature as the eternally pure would stop their ears against. We must hear 'body' in Foucault's discourse, not as a metonym for nature as opposed to culture, but rather as a term referring to no thing but standing in opposition to our desire for a sure and singular source for the truth of man.

Friedrich Nietzsche often used the term 'body' in a similar way. For Nietzsche, 'body' often functions as the hammer with which one shatters the idols of original or telic purity. I close this paper, then, with a quotation from *Daybreak*,[6] a quotation in which Nietzsche makes evident his departure from traditional quests for constancy, for a defeat of time.

> *The new fundamental feeling: our conclusive transitoriness.*—Formerly one sought the feeling of the grandeur of man by pointing to his divine *origin:* this has now become a forbidden way, for at its portal stands the ape, together with other gruesome beasts, grinning knowingly as if to say: no further in this direction! One therefore now tries the opposite direction: the way mankind is *going* shall serve as proof of his kinship with God. Alas this, too, is vain! At the end of this way stands the funeral urn of the *last* man and gravedigger (with the inscription '*nihil humani a me alienum puto*'). However high mankind may have evolved —and perhaps at the end it will stand even lower than at the beginning! —it cannot pass over into a higher order, as little as the ant and the earwig can at the end of its 'earthly course' rise up to kinship with God and eternal life. The becoming drags the has-been along behind it: why should an exception to this eternal spectacle be made on behalf of some little star or for any little species upon it! Away with such sentimentalities! (*ibid.*, sect. 49).

LADELLE McWHORTER
Northeast Missouri State University

[6] R. J. Hollingdale, trans. (New York: Cambridge, 1982).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

FOUCAULT'S CRITIQUE OF THE LIBERAL INDIVIDUAL*

IF liberalism, as a convenient abstraction, is political thinking that seeks to establish political values, structures, and institutions by discovering—not assigning—an inherent priority and privilege to pre-existing individual human beings and their spontaneous, self-directed activity, then Michel Foucault's investigations into the rationalities by which we have recently constituted ourselves as historical individuals seriously challenge the presumptuous arrogance with which liberalism continues to present itself as the theoretical basis for progressive social criticism and resistance to totalizing trends. If Foucault's insights are valid, liberalism's flaw is not that it has been inefficacious in pressing its agenda on behalf of individuals against modern tendencies that repress individuality and obstruct its potential, but that its rhetoric and its practices are themselves tragically and completely implicated in the burdensome network in which we find ourselves. His observations on the intertwinement of liberalism's individuality with its apparent opposite suggest that the current task for thinking and for action is not yet another attempted revivification and return to the individual, but instead is the rejection of, and resistance to, the individualities that we are, one crucial element of which is liberalism's programs and rhetoric.

Widespread individuality is a fairly recent idea and phenomenon. According to Foucault's analyses, the character of individuality did undergo a sharp, eruptive change at about the point in time which liberalism nostalgically recalls as its genesis, but this evolution is not the difference that liberalism's history expects.[1] By both accounts, the dominant history of liberalism and Foucault's countermemory, the distinctive sovereignty of the monarchical individual passed away; but they diverge on its replacement. For Foucault, there existed until the eighteenth century a formation of power in which only a few, those at the apex of formal, explicit structures, truly were individuals: the monarch was individuality's archetype, and individuality was greatest at the top of the political pyramid. Individuality was a rare opportunity and a privilege, a position of honor occupied

* To be presented in an APA symposium on the Thought of Michel Foucault, December 30. Judith Butler and Ladelle McWhorter will be co-symposiasts; see this JOURNAL, this issue, 601–607 and 608–614, respectively, for their contributions.

[1] This paper will draw primarily and heavily from Foucault's most extended analysis of the constitution of the modern forms of individuality, *Discipline and Punish*, Alan Sheridan, trans. (New York: Vintage Books, 1979).

0022-362X/89/8611/615–621 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

fully by the king and derivatively by his minions. The power of this style of individuality was exercised intermittently but with unignorable force; it was the occasional outburst of a lightning flash that seared its way into the undifferentiated multitude. This power had only one direction: it radiated from a focus. The writing associated with this individuality was the chronicles of monumental deeds and blood heritages which generated and preserved its distinctiveness from the mass. Its architecture was massive monoliths of force and impenetrability: the striking palace, the threatening dungeon, the inspiring cathedral. This power, in all its forms and symbols, supported and served pre-existing individuality; individuality was its cause and originating agent.

Liberalism would end this style of individuality by extending it, substituting for its infrequency its universality, melioristically replacing the arbitrariness of monarchical abuses and potential abuses with faith in the reason of each person. Rather than a single sun, which no matter how brilliant cannot light up the entirety of the social landscape, there would be an infinity of suns, illuminating every corner, so that decision and action would have no dark side, no obscurity or secrecy. Liberalism's project would not be to dissolve this form of social power, but to disperse it, to intensify it and its potential by granting each human being the status of individuality. It would make everyone a king.

Foucault's investigations into discourses, practices, and institutions from the late eighteenth century on do not, however, find liberalism's theoretical and prescriptive promise fulfilled in mundane reality. The dominant individuality from this epoch forward is not a spreading enfranchisement of the many into the previous style of the individuality of sovereignty, but is instead its inversion and alternation in every dimension. Contrary to liberalism's rhetoric of expanding the prerogatives and efficacy of the individual, this new individuality is liberalism's dark backside. It is clearly discontinuous with the dominant individuality that preceded it and unprecedented in its particular forms, but it does have its own ancestry, with roots that can be traced in the pastoral, governmental, and police mentalities of governing.[2] Unlike the style of ruling connected with sovereign individuality, which intervenes in the polity irregularly so that it maintains its power merely by the threat of its actions, the pastoral,

[2] For Foucault's analyses of the pastoral, the governmental, and the police, see "Omnes et Singulatim: Towards a Criticism of 'Political Reason'," *The Tanner Lectures on Human Values, Vol. II*, Sterling M. McMurrin, ed. (Salt Lake City: Utah UP, 1981), pp. 223–254; and "Governmentality," *Ideology and Consciousness*, VI (Autumn 1979): 5–21.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

governmental, and police modalities of ruling insist upon constant intervention and sustenance within the political body. Sovereign individuality holds its individuality in reserve, saving its strength for the explosive actions that then prevent that which it will not tolerate; its mode is repressive, through the capacity to extinguish the life of those who will never become individuals. The new form of individuality does not, however, threaten constantly, if quietly, to erupt to repress or erase life; it instead nurtures life by its continuous involvement in the details of society—it is positive and productive. It aims at making and encouraging life; its model is the good shepherd who will not forget a single sheep. This increase of life is achieved through mechanisms of individuation; it makes use of detailed knowledge of each individual and of the aggregate information collected through knowing individuals. Individuals are constituted in being known as individuals.

Every aspect of the old individuality is reversed. The chronicle of monumental deeds is replaced by the dossiers of the minutest behaviors of particular school children, patients, soldiers, criminals, inmates—written records which follow their lives and define them as individuals, which make them the individuals that they are, which make them individuals. The architecture of imposing external structures is exchanged for a problematization of useful internal spaces; buildings are no longer to strike awe so that the bulk of the population is held below the threshold of individuality. Instead, light, lines of sight, partitioning, and visibility are carefully calculated so that they may construct individuals; transfixed in the glare of radial visibility and lateral isolation, human beings have no choice but to become individuals in institutions and practices from the army to factories to schools to hospitals to reformatories, even to panoptically utopian prisons. Individuality is no longer an opportunity that lags as the political reforms of liberalism take their time to become effective; it is an imposed, inescapable necessity, a forced and enforced requirement. Attaining individuality is not graduating to a subjectivity that would exercise autonomy and spontaneity; in the institutions, discourses, and practices of the human sciences, individuals are constituted as the particular objects that have a dynamic of subjectivity. They are built so that they must be constantly in search of themselves and, ironically enough, so that they perpetually fail the criteria set for them and thus need ceaseless effort and re-examination, re-immersion in that which forms them.

The nexus and operations of power are also turned inside out; power is no longer the sovereign spasm that expends itself against its undifferentiated object and then is reflected back to show the superi-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ority of its individuality. Instead, power intensifies and reduplicates itself as it circulates endlessly through the nodes of its network. These points of articulation are the individuals formed in this network; as this power multiplies itself, these individuals are more and more clearly defined. Individuality becomes less an agential locus of origination than this power's vehicles, its instruments and means, the sites at which it is entrenched. This individuality is a product, an effect, not potential waiting to be emancipated. Punishment, for example, becomes correction, no longer exclusion or erasure, but irresistable inclusion and refinement of a new individuality. Individuality is greatest at the base of the social pyramid.

This power operates on individuals through measurement and observation without respite. It establishes mechanisms of unhesitating surveillance with pervasive standards of normality, so that each movement of every body is recorded, graded, and adjusted. It does not so much wear down its bodies into conformity as it elicits from them automatic and perfectly identical and repeatable actions. Whether in the context of working, learning, soldiering, or rehabilitating oneself, responses and individuals become homogeneous and interchangeable. Through examination and drilling, bodies and the souls which have been appended to them become docile: paradoxically, at the same time that their power and effectiveness has been enhanced, they have been made obedient and unhesitantly responsive, stripped of the capacity to be resistant by the manner in which they have been formed. They are the results and the possibilities of what Foucault calls the disciplinary apparatus.[3] The unit of social analysis is no longer the population as a single bulk entity—the wholesale control of the population—but the population as infinite aggregation—the retail management of bodies.

Ironically, according to Foucault's analyses, the individuals of discipline are not simply the warped, corrupted perversions into which the promise of liberalism has degenerated. Instead, the lineage of the disciplinary firmly precedes the programmatic ideology of liberalism. In fact, Foucault[4] understands liberalism as primarily reactive, not prospective. It was generated as a hesitating reply to the burgeoning

[3] For Foucault, the disciplinary is an analytical shorthand, a style or tendency of a form of power: although its practices do tend to coalesce into certain patterns, this is an emergent similarity, not an imposed univocity. It is not at all the product of any identifiable subjective intention, even that of history or class strategies. It is disparate but mutually supportive responses to differing local needs and conditions. Cf. *Discipline and Punish*, pp. 26/7, 138/9, 159, 224.

[4] "Foucault at the Collège de France II," *Philosophy and Social Criticism*, VIII (1981): 349–359.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

growth of the governmental and police mentalities, which in the seventeenth and eighteenth centuries took as axiomatic the endless expansion of government and its eventually pervasive infiltration into every sphere of human life for the sake of increasing life and the strength of the state. Liberalism was conceived, according to Foucault, as a theoretical stopgap to this administrative, informational, and manipulative steamroller, which had as its instruments, not only formal arms of governmental entities and agencies, but also such tools as private benevolent associations and professional organizations. Liberalism sought to limit the scope and the intrusiveness of the formal political state, marking off economic and private realms that would be immune from its invasions. Foucault's analyses decline to emphasize grand schemes and proclamations, to follow liberalism into the theoretical preserve that valorizes the individual as a politico-ontological privilege of the true kernel of humanity which stands behind and outside all social and political practices as the ultimate point of resistance to their hegemony. Instead, Foucault's emphasis on operative practices and discourses, the mundane historical reality of the individuals that have been our social products and effects, parodies by contrast and juxtaposition the serenity of that theoretical myopia with produced effects rather than theoretical desiderata. The irony of our recent individualities is the massive machinery of disciplinary individualities which almost silently flourished in the shadow of liberalism's vocal demands for liberty and autonomous individualism.

The harsher irony is not just that the prospective promise of liberalism's hopes for the individual were diverted into grotesquely opposite caricatures, but that liberalism itself is thoroughly entangled in that which vitiates it and renders it effectively irrelevant. Liberalism is not merely frustrated by the disciplinary apparatus of individualities alongside it; it quiescently participates in its own demise, thoroughly co-opted by their peaceful coexistence. The profound rights for individuals which were secured by great political documents and legal traditions were made insignificant by the complacent dependence of these formal institutions and processes on the expertise of the human sciences that legitimize the disciplines. This is most apparent in the decisions of a judicial system, which although formally radically insistent on the rights of the juridical individual, completely relies on, and defers to, the judgments made about individuals who have previously been made in the disciplinary archipelago or who are about to be made in it. The beautifully illuminated façade of liberal justice rests discreetly upon the obscured foundation of disciplinary machinations; the hard-won system of legal rights

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

and procedures for the juridical individual rides epiphenomenally on quieter, but more effective processes. By determining which cases enter the formal judicial channels and which are diverted to other dispositions, and by taking control of the length and style of punishment/treatment, the human sciences displace the significance of a judicial system that merely permits and ratifies. Liberalism's ideological game of society as the fair but difficult coordination of human monads into an effective but limited contract continues to dominate political discussion, but its expression is shadowed and rendered vacuous by the corporal technologies of discipline. The theoretical discourse of punishment, for example, remains at the level and presupposition of the contract: How is punishment to expel autochthonous individuals from the contract, rehabilitate them, and then restore them to the contract while still respecting their capacity for autonomy and independence? In fact, however, the institutions of punishment took a different direction—a problem not of justification but of efficiency: How are bodies and behaviors automatically to be formed and *re*formed into obedience?[5] The political issue becomes a social problem. In punishment, the clean system of liberalism depends on and uses dirtier disciplinary mechanisms; the ideal relies on more thorough techniques. The intensification of economic processes which marks the nineteenth and twentieth centuries is also more compatible with the disciplinary individuality on which it depends than with the liberal individuality whose pretenses it supports and makes possible. Egalitarian concerns and liberalism's worries about such theoretical issues as property give way to the needs of the productive system's need for useful, efficiently laboring bodies. We have two models of utopia, two narratives, the first of which we like to recall, the second we conveniently forget. The first is the political myth of the state of nature and the contract that corrects its decline. This generates the bourgeoise code of liberty, rights, and representative regimes. The second is the military and medical dream of comprehensive management, so that everything is included and utilized, so that nothing escapes (*ibid.*, pp. 146, 169). "The 'Enlightenment', which *discovered* the liberties, also *invented* the disciplines" (*ibid.*, p. 222).[6]

The individual as invoked by liberalism, therefore, is not an analytic fulcrum for critical and resistant thinking, but is instead, by its masking function that pretends that there is effectiveness to the

[5] *Discipline and Punish*, pp. 128/9.

[6] Even in its Roman inspirations, liberalism had this dualistic ambiguity, admiring both the citizen and the legion (*ibid.*, p. 146).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

liberal ideology, a pervasively compliant ideological element in the contemporary reality of individuality which we might well find intolerable. Despite its constant and vociferous protestations, it is serenely drawn into the influence of its negation. In our recent history, criticism in the name of the liberal against the disciplinary has consistently been swept, in its practical efforts, back once more into submission before the disciplinary. They are intractible partners in our present. Rigorous critical thinking, then, would not attempt once more to revitalize that which has repeatedly failed to be revitalized: the prison is a prime example in which a combination of the values of liberalism and the practices of the human sciences have failed to recover and enable the liberal persona supposedly buried beneath unfortunate personal and social circumstances, and have instead honed a fractured, recidivist, criminal individual to be managed. The dynamic that is this repeated failure is itself a component in the individualities that we make ourselves, continuously in need of more disciplinary and human sciences' attention for the sake of our liberal potential. Rather than fight again and again to preserve the dismal, futile hope of our liberal individuality, transgressive thinking would renounce our liberal individuality, declining to be caught in its vortex. Constantly interrogating the rhetoric of individuality with the practices of individuation with which it is complicitous, it would lightly suspend our habit of commitment to this language in the face of the mocking effects and products associated with it. Tracing the elusive, fragmentary connections between the rhetoric and its reality, it would exhaust the force of any claim that we must maintain our respect for our liberal individuality, if we are to hold our respect for ourselves; it would generate a calculated indifference that breaks the spell of the liberal over us, so that we would be less likely to slide immediately and expectantly back into its demands. We would exit our liberal individuality, not by simply transcending or rejecting it, but by working through it; within it but not casually partaking of it, we might break its seductive temptations. For once, we would experience, not an emancipation and actualization *of* our individuality, but an emancipation *from* our individuality.

DAVID F. GRUBER

Northeast Missouri State University

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

WITTGENSTEIN'S LATER PHILOSOPHY OF MIND: SENSATION, PRIVACY, AND INTENTION*

WHAT follows is by way of a long abstract of a sequel to another paper.[1] I want to point, very schematically, to a connection, well worth further study, between the kind of doubt about the possibility of private language which *Investigations* §258 should be seen as suggesting and an idea about our knowledge of our own intentional states—beliefs, desires, intentions, etc.—for which I have also claimed a Wittgensteinian origin in recent papers.[2] But I shall try to presuppose no familiarity with those discussions.

The basic ideas in *Investigations* §258 are, first, that the "statements" in a private language of sensation would be contentful, only if there were a substantial distinction between what it seemed right to the user of a private language—a linguist—to say or think, and what really was right; and, second, that there can be no such distinction. The connection of this thought with the nature of knowledge of one's own intentions turns on the simple point that a linguist would need, one way or another, to generate the semantics of the private language out of his own intentions. Roughly: correct identification of a sensation would involve applying the identifying symbol to a sensation of the type with which the linguist intended that it should be associated when he did whatever he did do to establish that symbol's meaning. The supporter of private language has an obligation, therefore, to ensure that there is a coherent notion of intention of the kind which the linguist needs; specifically, a notion of intention

* To be presented in an APA symposium on the Thought of Wittgenstein, December 30. Warren Goldfarb will be co-symposiast and John McDowell will comment; see this JOURNAL, this issue, 635–642 and 643/4, respectively, for their contributions. The material abstracted here was composed for the conference on Metaphysics and Epistemology held at the University of Queensland in May 1989. I am grateful to those who participated for helpful discussion, and especially to Andre Gallois. Thanks also to Bob Hale for comments on a draft of the paper presented there.

[1] "Does *Philosophical Investigations* I, §§258–60 Suggest a Cogent Argument against Private Language?" in P. Pettit and J. McDowell, eds., *Subject, Thought and Context* (New York: Oxford, 1986), pp. 209–266.

[2] "On Making Up One's Mind: Wittgenstein on Intention," in *Logic, Philosophy of Science and Epistemology, Proceedings of the XIth International Wittgenstein Symposium,* P. Weingartner and Gerhard Schurz, eds. (Vienna: Holder-Pickler-Tempsky, 1987), pp. 391–404; "Wittgenstein's Rule-Following Considerations and the Central Project of Theoretical Linguistics," in Alexander George, ed., *Reflections on Chomsky* (New York: Blackwell, 1989), pp. 233–264; "Critical Study of Colin McGinn's *Wittgenstein on Meaning,*" in *Mind,* XCVIII (1989):289–305.

0022-362X/89/8611/622–634
© 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

such that by forming one on a "baptismal" occasion, as it were, I can—even in circumstances of privacy—create facts about the proper description of sensations that await me in the future. That is an obligation which, on any view about knowledge of one's own intentional states along the broad lines I am going to canvass, cannot be discharged.

I. THE SEEMS RIGHT/ IS RIGHT DISTINCTION

According to the broadly Cartesian conception of the mental which Wittgenstein is examining, the truth of any of a linguist's judgments concerning his sensations will originate somewhat in the fashion of the truth of a contingent report of observation: it will be a matter of depiction of an appropriate—albeit interior—truth-conferring state of affairs, something manifest to consciousness but distinct from the judging, whose content is precisely that such a state of affairs obtains. The existence of a contrast between what seems right to the linguist and what is right is simply a requirement of this comparison: something demanded if it is to be appropriate to think of matters in terms of a possible fit, or congruence, between the linguist's judgments concerning his sensations and the character which those sensations actually have.

It is no cause for concern about the propriety of this demand that we are concerned with judgments about states of which the linguist is necessarily aware. For the linguist's—and our ordinary—conception of sensations allows for their occurrence in subjects, like animals or very young children, who lack the conceptual wherewithal to judge and classify them. So the point stands: if judgments about sensation are to be viewed on the Cartesian observational model, we should conceive of their correctness, when they are correct, as rendered by something distinct from the judgment—the character of the sensation, the fact that it falls under the concept under which it is thereby judged to fall.

If a seems right/ is right distinction is necessary, what would suffice to make it out? The implied answer of §258 is: some independent "criterion" of correctness—some method whereby the linguist could further appraise judgments that initially strike him as correct. But the privacy of the subject matter, and of the concepts under which the linguist is putatively bringing it, necessarily pre-empt, Wittgenstein seems to suggest, the possibility of any such criterion. So, he is tempted to conclude, no content can be established for the needed idea of congruence between the linguist's successive convictions and the putative states of affairs they concern. And, hence, "here we can't talk about 'right'."

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

There are two salient lines of defense open to the friend of privacy. One of them[3] contests whether the linguist does irremediably lack all means for a principled appraisal of his own judgments. The suggestion is that, by constructing sufficient of a system—theory—around his judgments, the linguist could establish conditions of acceptability which related not just to the judgments' initial appeal but to principled holistic constraints. And he could thereby have, in certain circumstances, compelling reason to overturn initially appealing judgments that he would otherwise have accepted. That would surely be to possess a "criterion of correctness" in the sense material to §258.

I investigated this proposal in some detail in an earlier paper (cf. fn. 1). Its prospects are, in my view, extremely unpromising. But our present business is to canvass an argument to the effect that the friend of privacy has no alternative course.

The second response questions whether, in order for the linguist to be entitled to the needed distinction between his judgments and the states of affairs to which they are supposed to correspond, a further criterion is necessary for their acceptability beyond their commanding his initial assent. On the contrary: difficulty in making out a contrast between what seems right to the linguist and what is right is exactly to be expected if, as we naturally suppose, one's judgments about one's own sensations are infallible. In that case, so far from wanting a criterion for the correctness of his judgments, the linguist has such a decisive criterion in the very phenomenon of judgment itself. No other criterion is necessary or, if it is intended that it should dominate and, on occasion, correct the linguist's impressions, possible.

A seems right/ is right distinction was demanded in order to vindicate the idea that the linguist has a genuine class of recognitional judgments in play, answerable to a genuine range of situations of which it is possible—if only for the linguist—to have a reasonably definite conception. The second response brings out that no fast-track demonstration that these ideas have no content can be based purely on the impossibility of ascertainable contrasts between what the linguist is/was inclined to think and what is/was actually so. For if there can be such a thing as infallible judgment, of any subject matter, then there will be both a genuine distinction between the judging and the states of affairs judged and no possibility of that kind of ascertainable contrast. Pleading infallibility, as it were, does not,

[3] See Ross Harrison, *On What There Must Be* (New York: Oxford, 1974), chs. 3, 6; note esp. p. 161; and Simon Blackburn, "The Individual Strikes Back," *Synthese*, LVIII (1984): 281–301, concluding pages.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

of course, establish the needed distinction,[4] but it points up what might seem to be a lacuna in Wittgenstein's reasons for thinking that it cannot exist.

II. INFALLIBILITY

How, though, can judgment, even about so intimate a matter as one's own sensations, ever be infallible? It is true, of course, that there are many ways in which one cannot, apparently, go wrong in making such judgments. There is, for instance, no distinction between a sensation's appearance, as it were, and its reality, so no ways in which appearances can deceive; there are no analogues, for sensation, of bad lighting, illusions of perspective, and background noise. There is also no analogue of the distinction between how a perceived object seems to a particular subject and how it publicly seems; for the way a sensation feels to someone is the way it feels, *simpliciter*. But such considerations, though limiting the range of possible sources of error, do not establish infallibility. Judging a sensation properly, as the linguist conceives it, is a matter of responding with an appropriate concept to a token of a type of interior state of affairs which could have been instantiated in his consciousness even had he lacked the capacity to make judgments about it altogether. So it should be viewed as a substantial cognitive accomplishment—a simple but genuine feat of phenomenological taxonomy. As such, it calls, one would imagine, for wheels to turn within one's cognitive apparatus—for presumably causally based classificatory mechanisms to fire. Why can they not misfire? Is it even intelligible that they might somehow be proof against misfire?

In §271 Wittgenstein has his interlocutor say,

> Imagine a person whose memory could not retain *what* the word 'pain' meant—so that he constantly called different things by that name—but nevertheless used the word in a way fitting in with the usual symptoms and presuppositions of pain.

Wittgenstein continues *in propria voce,*

> —In short, he uses it as we all do. Here I should like to say: a wheel that can be turned though nothing else moves with it, is not part of the mechanism.

Here the "idle wheel" is that putative aspect of the competent ascription of pain to oneself and others which is supposed to involve

[4] Rather, the propriety of regarding a class of judgments as genuinely infallible, as opposed to merely incorrigible, depends precisely on making out the distinction. Cf. my *Wittgenstein on the Foundations of Mathematics* (Cambridge: Harvard, 1980), pp. 356/7.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the proper classification of an interior object, and hence to depend on success in remembering how the interior object has to be if it is properly to be classified as 'pain'. Wittgenstein's suggestion is that we should be doubtful whether any such interior classificatory competence is involved if somebody could lack it yet apparently function in all respects normally in his use of the word. Let that pass, however; does not the example at least force the linguist to acknowledge one kind of fallibility? Clearly the sincere use of a private language cannot be infallible. For if the linguist can establish private meanings, he can also lose track of them and use symbols in ways which diverge from the requirements of the meanings which he himself established.

That is a relatively superficial point. But it takes us close to a deeper one. At *Investigations* II, xi (p. 272), Wittgenstein writes:

> Always get rid of the idea of the private object in this way: assume that it constantly changes, but that you do not notice the change because your memory constantly deceives you.

My interest here is not in the efficacy of this thought against the "private object," but in the extent of the memory deception that Wittgenstein has in mind. Consider this possible parallel: you watch a colored light that varies between red, green and blue, and report its color at specified intervals; you do so correctly, but at each stage you misremember both how the light was at the last stage and how you classified it, in such a way that it seems to you that there has been no variation in color at all and that you have all along been describing the colors correctly. Here there is no question about your continuing competence with color concepts—for all your judgments of present color are correct. Your failings belong entirely to your memory of facts.

It is different on the following possible reading of the passage. Now—when it comes, say, to reporting the light's color for the second time—you correctly remember your previous classification but misremember how the light looked, taking it that there has been no change. Finding yourself now inclined to a different classification, you may be inclined to disown the previous one; but equally, its precedent may stifle your inclination to what would in fact be the correct description of the present color of the light. In this way, a failure of factual memory may destabilize your grasp of color concepts.

It is important to see that such destabilization need not merely be verbal. Someone learning a second language might occasionally muddle the names of colors, calling purple 'amarillo', for instance, or orange 'rojo'. The beliefs about colors which, thus incompetently,

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the subject tried to express could and probably would, however, be perfectly correct. It is not merely an analogue of this kind of muddle which the linguist has to acknowledge to be possible in view of the fallibility of memory. Consider instead the case of a child learning to classify colors in English as his first language. Here it is not a matter of mapping a vocabulary onto an antecedently grasped set of concepts; rather, the boundaries of the concepts are set by the linguistic practice. And mastery of the vocabulary is not an all-or-nothing matter; various degrees of grasp are possible. There will be stages at which the child is acquainted with some but not all familiar color terms; and, more significantly, there may be stages at which his grasp of the use of particular color terms is partial. He may, for instance, reliably classify blue things as 'blue', and red things as 'red', but be inclined, until corrected, to classify most shades of purple as 'blue' and some—but perhaps not all—of the remainder as 'red'. Here it would be improper to say that his errors are merely verbal; for if they are verbal, then he is misusing the sentence, 'This is blue', to express true beliefs, ergo beliefs that he has the conceptual resources to entertain, about the colors of purple things. But any relevant true beliefs are to the effect that the objects in question are indeed purple; so to regard his errors as verbal would commit us to ascribing the concept of purple to him which, since grasping it involves mastery of linguistic boundaries, is precisely what he lacks.

There are degrees of grasp, not just of color vocabulary, but of color concepts, and, correlatively, degrees of confusion. Normally, people move through such intermediate degrees of grasp to attain a standard competence in color judgment which stays with them for the rest of their lives. But the durability of the competence is a contingency, like the durability of the capacity to balance on a bicycle. A kind of recidivism is evidently possible, and might have been quite usual; human existence supplies many examples, after all, of gradual loss of capacities that, in the earlier part of our lives, we may have worked hard to acquire.

The linguist, then, who would conceive of his ability to classify his sensations as an inward-looking recognitional skill, must recognize two potential kinds of error which the fallibility of his memory may engineer. First, he may suffer *linguistic amnesia*—forget which expressions respectively denote which kinds of sensations—although retaining a grasp of the various sensation kinds which, in the language as originally instituted, he had determined the use of the various classificatory terms to signal. Second, he may suffer from a kind of *conceptual recidivism* parallel to that of someone who, having once mastered the system of color concepts reflected in the

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

standard vocabulary in English, slips back into, e.g., classifying a large class of shades of purple with the blues. The linguist may slip, that is, into classificatory practices that, judged in the light of his own previous practice, would reflect a merely partial grasp of the concepts that were constituted therein. And this may happen while his dominant intention is to conserve his original classificatory scheme.

Conceptual recidivism is, evidently, merely a special case of a more general possibility: departures from the system of concepts enshrined in the linguist's original practice may take all kinds of, perhaps unprecedented, directions. What is important, for our purposes, is that something other than notational variation be involved, and that the linguist's dominant intentions remain conservative.

Now, there is a question whether either kind of possibility makes space straightforwardly for the linguist's holding a false belief about the character of an occurrent sensation. For the possibility of linguistic amnesia is only the possibility that he forms a true belief that he then forgets how to express; and, in a case of conceptual recidivism, his grasp is in doubt of the concepts with which, in ascribing a false belief to him, we should apparently be crediting him. Plausibly, I can only believe falsely that something is blue, if I know what it is for something to be blue and falsely believe that the particular object is that way.

There is scope for some debate about the last claim. It does not, for instance, seem a major solecism to describe someone who is confused about the distinction between leopards and cheetahs as believing of the cheetah in the cage in front of him that it is a leopard. But we can, I think, forego this debate. For if conceptual recidivism and its ilk are possibilities that the linguist must acknowledge, then he must at least allow the possibility of *second-order error:* the possibility of a situation in which he falsely believes that he has a true classificatory belief about an occurrent sensation, even if he is no longer competent to identify the content of the belief that he falsely takes to be true—since his error results from his having lost hold of concepts of which, in identifying that content, he would be exercising a continuing grasp. True, if matters are described in this more conservative way—and if loss of grip on his own vocabulary and conceptual recidivism and its ilk exhaust the pitfalls to which the linguist's attempts to annotate his sensations may succumb—then, whenever the linguist does manage a classificatory belief of determinate content, that belief will be true. The proper conclusion, however, is not that the linguist is infallible in his attempts to classify his sensations, but that infallibility requires more than that the classificatory beliefs that he actually succeeds in forming be error-proof: it

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

requires, in addition, that there be no possibility of his being deluded that he has such a true belief when in fact he does not.

III. PRIVACY AND INTENTION

Infallibility is not to be had. But if the sort of second-order error which attends the possibility of conceptual recidivism and similar lapses is the worst the linguist has to reckon with—and I am supposing it is—does that put obstacles in the way of a successful argument against private language? Would it be more straightforward to design such an argument, if we had forced the linguist to face the possibility of simple misclassification—false first-order beliefs?

I do not think so. If the linguist makes such a second-order error, there will be a distinguished belief which, though he has unintentionally drifted away from the concepts necessary to grasp its content, he nevertheless takes himself to hold truly—a belief expressed by the private sentence by which he attempts to represent his judgment to himself, but expressed by that sentence courtesy of the private language's original semantics which it remains his dominant intention to conserve. And this distinguished belief may actually be false in the situation in question. (Cf. the color recidivist's use of 'This is blue' when confronted with something purple.) Hence, exactly as there would be if first-order error were involved, there is a need to conceive of the source of the truth value of the distinguished belief as residing in factors that are independent of—and may actually outrun—the linguist's capacity for competent judgment. So the challenge remains to explain what these factors are.

Our concern is with any response to this challenge of a nonoperationalist temper—any response that does not regard it as necessary, in order for sense to attach to the idea of the linguist's being in error, that his errors be detectable. If he errs undetectably, what makes it the case, e.g., that the relevant distinguished belief, which he falsely believes that he holds truly, is in fact false? There is only one possible answer. The concepts of which the linguist now has at best an imperfect grasp have to be thought of as constituted by certain of his prior linguistic intentions: either the "baptismal" intentions that were operative when he gave himself the relevant private ostensive definitions, or—since the way the private language comes into being is strictly an irrelevance—intentions which previously informed his mature practice of the language and with which he intends his present practice to be consonant. And his error can consist only in his mistaking what constitutes compliance with those intentions on the occasion in question—a mistake originating not in any sort of misapprehension of the prevailing circumstances, for which the Cartesian epistemology of sensation provides no scope, but in loss of grip

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

on the intentions themselves. Whence the connection of the private language issue with our knowledge of our own intentions and of our intentional states in general.

How, broadly, is it possible to have such knowledge? Two paradigms are liable to blinker our response to this question: knowledge by inference and knowledge by observation. But the idea that knowledge of one's own intentional states is more than quite exceptionally arrived at inferentially misrepresents both their phenomenology and their "grammar." We do not in general have any account to offer of the grounds on which we take ourselves to have the beliefs, intentions, etc., that we do, and it is normally quite out of place to expect such an account. Intentional states are *avowable:* they are subject to groundless, authoritative self-ascription, and it belongs to their essence that they are so.

The Cartesian conception of knowledge of one's own intentional states opts for the other, observational paradigm: the intentional state is, typically, simply salient to consciousness, and directly recognized by the subject for what it is. This idea is a major target of Wittgenstein's destructive effort in the *Investigations*. A currently fashionable sort of reason for discontent with it, emanating from the writings of Hilary Putnam, Tyler Burge, and others,[5] would reflect that the content of intentional states in general is determined as a function of factors—for instance, aspects of the physical and/or social environment—which lie out with the subject's consciousness; so the identity of an intentional state cannot in general be ascertained merely by scrutiny of elements that lie within consciousness.[6]

[5] The *locus classicus* is Putnam's "The Meaning of 'Meaning'," in his *Mind, Language and Reality* (New York: Cambridge, 1975), pp. 215–271. See also his "Is Semantics Possible?" *ibid.*, pp. 139–152. Burge's ideas concerning the contribution of social considerations are introduced in his "Individualism and the Mental," in *Midwest Studies in Philosophy,* IV (1979): 73–121; see also his "Other Bodies," in A. Woodfield, ed., *Thought and Object* (New York: Oxford, 1982), pp. 97–120, and "Two Thought Experiments Reviewed," *Notre Dame Journal of Formal Logic,* XXIII (1982): 284–293.

[6] Two points about this train of thought. First, its thrust is not that second-order beliefs become additionally fallible on an externalist view of content; on the contrary, if the content of the belief that *P* is in part externally determined, so—and in exactly the same way—is that of the belief that I hold the belief that *P*. Externalism poses no obstacle to the idea that, if I believe that I believe that *P*, then I do believe that *P*. The question is, rather, how I am to tell introspectively what second-order belief I hold. Second, the train of thought is, of course, utterly unconvincing as formulated: it is rather as if someone were to argue that it was impossible to recognize a friend in the street, since the identity of the approaching person would depend on considerations of history and origin which could not be salient to contemporary observation. There is much more to say. For disbelieving responses to the train of thought, see Donald Davidson, "Knowing One's Own Mind," *Proceedings and Addresses of the American Philosophical Association,* LX (1987):

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Wittgenstein's own objections to the observational model are different, however. In part, they are once again phenomenological, though there is no space here to review any of the details. But perhaps the most decisive difficulty that his writings disclose derives from the requirement the model imposes that intentional states can be fully determinate objects of inner contemplation before they issue in anything outward—and hence that their connection with events, and especially performances, lying in the future can only be causal. For, on the contrary, the ascription of intentional states to a subject is answerable to what he goes on to say and do in the broadly identificatory fashion in which the ascription of dispositions and capacities is so answerable. Indeed, an ascription of a specific set of beliefs, desires, and intentions to a subject can assume sufficient detail to ensure that the question of its justifiability in the light of his subsequent performance presupposes no knowledge of whatever causal connections there may standardly be between intentional states and behavior, but is a priori.

One of the most basic of the philosophical puzzles about intentional states is that they seem to straddle two paradigms: the paradigm of sensation and other avowable phenomena of consciousness, on the one hand, and, on the other, the paradigm of psychological characteristics that, like irritability or modesty, are properly conceived as dispositional and give rise to no phenomenon of avowal. It appears that neither an epistemology of observation—of pure introspection—nor one of inference can be harmonized with all their aspects. Now, when each of the horns of a dilemma seems unplayable, it is often good policy to look for a common assumption. And there is, of course, such an assumption here. It is the assumption that there has to be a substantial epistemology of intentional states, a mode of cognitive access to those states which is distinctively available to their subject and which is somehow able to measure up to the epistemic security with which sincere avowals of intentional states are standardly credited. But then it seems that the only relevant possibilities—since one does not know a priori of one's own beliefs, desires, etc.—are observation and inference, and neither seems to be at the service of a satisfying account. The conclusion would therefore seem suggested that knowing of one's own beliefs, desires, and intentions

---

441–458; and Burge, "Individualism and Self-Knowledge," this JOURNAL, LXXXV, 11(November 1988): 649–663. For defense of the train of thought as a weapon against the observational model, see Paul Boghossian's "Content and Self-knowledge," forthcoming in *Philosophical Topics*.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

is not really a matter of "access to"—being in cognitive touch with —a state of affairs at all.

How can this thought be made clear sense of? I strongly suspect that there is only one viable strategy for doing so. That is to recognize that the authority standardly granted to a subject's own beliefs, or expressed avowals, about his intentional states is a *constitutive principle:* something that is not a by-product of the nature of those states, and an associated epistemologically privileged relation in which the subject stands to them, but enters primitively into the conditions of identification of what a subject believes, hopes, and intends.

There are various, more precise possible accounts and comparisons by means of which one might attempt to render this idea more concrete,[7] but I have no space to attempt to take the task on here, and we need, in any case, only a very general sketch for our present purpose. The language game of intentional states, so conceived, would function roughly like this. A subject's sincere dispositions of avowal—or, better, his beliefs—concerning his intentional states would stand by default, unless there was positive reason to reject them. Such reason might be provided by the inability of the intentional system so determined to rationalize his behavior satisfyingly, together with the availability of an alternative system, generated by discarding the suspect ascriptions, which fared better. Or it might be provided by internal constraints of harmony and intelligibility working within the system as a whole. Since the telos, in the most general terms, of the practice of ascribing intentional states to oneself and others is mutual understanding, the success of a language game that worked this way would depend on certain deep contingencies. It would depend, for instance, on the contingency that taking the self-conceptions of others seriously, in the sense involved in crediting their beliefs about their intentional states, as expressed in their avowals, with authority, will almost always tend to result in an overall picture of their psychology which is more illuminating—as it happens, enormously more illuminating—than anything which might be gleaned by respecting all the data except the subject's self-testimony. And that in turn rests on the contingency that we are, each of us, ceaselessly but—on the proposed conception—subcognitively moved to opinions concerning our own intentional states which will indeed give good service to others in their attempt to understand us. Thus, we do not cognitively interact with states of affairs which

[7] See my "Wittgenstein's Rule-Following Considerations . . .," pp. 246–254; and my "Moral Values, Projection and Secondary Qualities," *Proceedings of the Aristotelian Society,* supp. vol. LXII (1988): 1–26.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

confer truth upon our opinions concerning our own intentional states; rather, we are inundated, day be day, with opinions for which truth is the default position, as it were. They count as true provided that we hold them and that no good purpose is served, in another's quest to find us intelligible, by rejecting them.[8]

Sketchy as it is, some of the effects of applying this austere picture to intention itself are foreseeable and striking. That I should know what I currently intend becomes the default position. But reflect that the picture, if good at all, ought also to extend to knowledge of our former intentions or—what comes to the same thing, barring mistakes about other matters—knowledge of what currently would comply with them. For making sense of others' behavior depends as much upon knowledge of their former as of their present intentional states; and, once again, our practice is to give avowals of former intentions, and judgments about what courses of action now comply with them, the same kind of default authority.

If anything like this is correct, then the linguist is in serious difficulty. We saw that the conception of a private language as founded upon inward-looking classificatory competences demanded acknowledgment of the kind of fallibility illustrated by conceptual recidivism. And that in turn demanded the possibility of a split between the classification of an occurrent sensation that accords with the linguist's baptismal or erstwhile mature intentions, and the linguist's actual classificatory performance in response to that sensation. The linguist's baptismal, or mature, intentions have to be empowered to constitute substantial truth conditions for beliefs that he has not yet formed about sensations that have not yet occurred; and they have to be able to do this in total independence of what, if any, beliefs he actually does form when the sensations occur. So we have to conceive of facts about what does or does not conform to those of the linguist's prior intentions which constituted the linguistico-conceptual apparatus that he is seeking to deploy, as settled quite independently of his own subsequent responses and judgments.

But no such conception can be accessible on the austere view of intentional states which was sketched. Simply: there is nothing for an intention, conceived as determining subsequent conformity and nonconformity to it autonomously and independently of its author's

[8] To avoid misunderstanding, note that the austere picture proposes a (default) sufficient condition for a subject's being in a particular intentional state. There is no suggestion that the same condition is also necessary. So the picture need not involve that there is always a regress of beliefs whenever anything is believed; or that unselfconscious creatures, which lack the conceptual resources for second-order beliefs, are incapable of believing altogether.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

judgments on the matter, to be. What the linguist needs is a platonized intention, an interior mental state sustaining objective relations of fit to prospective courses of action rather as a rule, platonically conceived, determines its own proper applications indefinitely far down the tracks, as it were—leaving the human subject, in both cases, only with the task of following through a predeterminate commitment. That is mythology. And the austere view has no place for it. With the austere view in place, the linguist is bereft of the means to give the needed constitutive account—to explain what, independently of his judgment, constitutes the truth of a thought expressed in the private language.

The reason the austere view does not generate a similar result in the case of public language is that keeping a grip on one's own prior linguistico-conceptual intentions does not, in contrast with the private case, suffice for truth; for one may be encumbered with the wrong intentions—misunderstand the content of the judgment at issue—or fall victim to perceptual or other kinds of cognitive error. It is the linguist's peculiar predicament, by contrast, that a failure to keep track of the requirements of his previous baptismal or mature intentions is the only way he can fall into error. Yet error—if only second-order error—has to be possible, if a genuine exercise of concepts is involved. The result is that the austere view of intention, by granting default correctness to all the linguist's private "judgments," makes it impossible for any of them to be correct—and hence undercuts the picture of a domain of independently constituted states of affairs which those "judgments" were supposed to fit.

The only possible recourse for the friend of privacy is to provide some account whereby default correctness might on occasion be overridden—something parallel to the way whereby explanatory constraints, operated by third parties, may dominate a subject's sincere avowals. But that is to say that any hope for the linguist must depend upon his turning explanatory theorist with respect to his own basic classificatory responses—seeking to devise criteria whereby they may be appraised. I do not believe[9] that there is really any hope in that direction. But, in any case, if the argument I have adumbrated succeeds in detail, no purely constitutive conception of the seems right/is right distinction can be sustained.

CRISPIN WRIGHT

Universities of St. Andrews and Michigan

[9] For the reasons elaborated in my "Does Philosophical Investigations . . . ."

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

WITTGENSTEIN, MIND, AND SCIENTISM*

LUDWIG WITTGENSTEIN'S treatments in the *Philosophical Investigations*[1] of the intentional mental notions are evidently meant to persuade us that, in some sense, understanding, believing, remembering, thinking, and the like are not particular or definite states or processes; or (if this is to say anything different) that there are no particular states or processes that constitute the understanding, remembering, etc. Such a dark point desperately needs clarification, if it is not to deny the undeniable; for surely we may speak of a *state* of understanding, or of thought *processes*. Now, Wittgenstein sometimes puts the point more specifically, speaking of "a state of a mental apparatus" (§149), or, with respect to understanding, of "a state which is the *source* of the correct use" (§146), as that which he wishes to deny. These denials may sound somewhat less paradoxical, but they are equally in need of elaboration.

However Wittgenstein's point is clarified, though, there is an objection that will inevitably be made. This objection asks: Is Wittgenstein here not usurping the place of empirical inquiry? Is it not possible that empirical science—neurophysiology, in particular—will find specific states and processes that will fill the bill, as far as understanding, believing, remembering, etc., are concerned? To preclude such an outcome now is just to claim that we shall never obtain certain sorts of results in future neurophysiology. But then Wittgenstein is simply making a bet on the future course of science; or else he is denying a priori that certain empirical investigations could be successful. Neither alternative is attractive.

I mean this *scientific objection,* as I shall call it, to be narrowly based, on envisaged possibilities of results in neuroscience; it is not the charge that Wittgenstein fails to leave room for the discoveries of psychology, cognitive science, or some as yet unknown science of the mind. Rather, the scientific objection exploits the unarguably scien-

* To be presented in an APA symposium on the Thought of Wittgenstein, December 30. Crispin Wright will be co-symposiast and John McDowell will comment; see this JOURNAL, this issue, 622–634 and 643/4, respectively, for their contributions.

I am indebted to Steven Affeldt, Louise Antony, Burton Dreben, Edward Minar, Angel Oquendo, and Thomas Ricketts for helpful discussions and suggestions.

[1] G. E. M. Anscombe, trans. (New York: Macmillan, 1953); passages will be cited by section numbers standing alone.

0022-362X/89/8611/635–642 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tific status of neurophysiology, and the independence from psychology of the conceptions of state and process that it employs. As a result, addressing the objection should help to elicit clarification of Wittgenstein's dark point about states or processes.

A common view of Wittgenstein takes his response to such an objection to proceed by noting that a "hidden" physiological state could not reflect central conceptual features of our notion of understanding (or believing, thinking, etc.). No such state would have conceptual links to correct responses and the other criteria of understanding; no state, in short, has the "grammar" of understanding. Hence, Wittgenstein is taken to urge, understanding cannot be conceptually constituted by any such state.

Although there are passages in the *Investigations* that suggest this response, to my mind it fails to capture Wittgenstein's thought. The response requires a general and sharp distinction between conceptual and empirical, between criterion and result of investigation, which I do not believe Wittgenstein accepts (see §79, last paragraph). Moreover, the response leaves it open that understanding, although not "conceptually" constituted by a neural state, might be "empirically" constituted by one.[2] This underreads Wittgenstein's conclusion. Thus, the suggested response does too little. Yet it is not totally off the mark, for Wittgenstein does think that the identification of any state of understanding would have to bear some conceptual burden. His conclusion will be that, in some sense, "nothing is hidden." But this is meant to result from detailed considerations of the notion of understanding, not to be assumed more or less as a premise, as an unargued feature of his methodology.

The notion that Wittgenstein dismisses or ignores questions of "empirical" identity also makes him look simply antiscientific. Indeed, Wittgenstein has been accused of being dogmatically hostile to projects aimed at scientific explanation. Now, it is true that Wittgenstein was distrustful of the claims of science and its role in contemporary culture (see, e.g., the foreword to the *Philosophical Remarks*[3]). It is central to Wittgenstein's teaching that a felt need for explanation in philosophy must be scrutinized; for, what it is, exactly, that

[2] The terminology comes from Paul Horwich, "Critical Notice: Kripke: *Wittgenstein on Rules and Private Language*," *Philosophy of Science*, LI, 1 (March 1984), p. 168.

[3] Rush Rees, ed., Raymond Hargreaves and Roger White, trans. (New York: Blackwell, 1975).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

wants explanation at all may only become clear through such an investigation. All this does not make him antiscientific. It does make him antiscientistic, against the smug and unexamined assurance that what wants explanation is obvious, and that scientific tools are immediately applicable. For Wittgenstein, scientism is just as misguidedly metaphysical as traditional, more transparently a prioristic, approaches. An immediate inclination to look to science for answers can in fact be an expression of a philosophical picture.

This point is made in §158, the one explicit appearance in the *Investigations* of something like the scientific objection. Here Wittgenstein suggests that, although the objection says that the scientific investigation may come out either way, it is only a scientific hypothesis or conjecture that such-and-such a process or state will be found, at bottom the objector is being moved by an a priori demand that things must turn out a certain way. The claim of a modest empiricism is mere lip service.

With respect to understanding, the point might be phrased thus. In our practices of ascribing understanding, there is an enormous variety of considerations that can enter, and there is a dependence on context impossible to subsume under general rules. It is just this that can drive one to think there must be some final item, behind the phenomena, which grounds the ascription, which is what the ascription is an ascription *of;* without this, the ascriptions would be arbitrary (cf. §146). If it is at such a juncture that the scientific objection is voiced, then the roots of the objection in an a priori requirement, that something unitary and definite ground the ascription, are evident. The objection thus becomes just a restatement of a picture of what has to be there. What is operative at bottom are a priori demands on the mental notions.

Such demands are the focus of Wittgenstein's investigations of mental concepts. If these are successful, we shall lose the idea of the specific particular process or state that is the understanding, along with the idea that we need such a thing. In this way, Wittgenstein's investigations are to have the effect of making entirely obscure what could be scientifically discoverable which would *be* the mental state or process of understanding—since, if you will, there is nothing for it to be, no role for it to play.

In Wittgenstein's treatment of understanding, I see two strands, perhaps separated only at the cost of some distortion. The first concerns how the state is conceived to operate. The idea under scrutiny is that the presumed state of understanding is the source

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

from which flow the manifestations of understanding. Wittgenstein asks whether we have any conception of the "flowing from," any conception of what we are asking of the state when we take it as issuing in, determining, fixing, what counts as a manifestation of the understanding. The second strand concerns the range of things the particular state or process is supposed to do, and involves examining the topography of understanding and manifestations of understanding in varieties of cases. It is here that what Wittgenstein calls a "grammatical investigation" figures.

Let me start with the first strand. The definite state of understanding that Wittgenstein wants to deny is meant to be the state that, in and of itself, in some way determines everything that counts as a manifestation of that understanding. Wittgenstein argues that this notion, and in general any notion of an "internal standard," is confused. It is based on a misconstrual of the models we have, in which pictures, formulas, samples, and so on, do issue in or determine the characterizations of items. It is to point out this misconstrual that he uses, several times over, a third-man style argument: with respect to samples in §74, signposts in §86, mental pictures in §139, and rules in §198. External standards do give us the means to characterize objects or to generate a sequence; but they do so because we have the practices and abilities to use them. The internal standard is supposed to give us the means to identify items, but without practices and abilities, for the internal standard is also meant to operate by itself, in a self-sufficient manner. (If it were not, it would be otiose: Why not settle for practices and abilities themselves?) In short, the internal, mental determinant is imagined on the model of external standards, but in the absence of the surroundings that make the standards into standards. This move is made in construing understanding, for the notion of the self-sufficient state is just the notion of an "internal standard" that legislates all the individual uses. Wittgenstein's point is to expose its incoherence.

Here, however, the scientific objector has a retort. Once physical science is in play, causal notions become available. Thus, the incoherence is avoided; what is being imagined as being in principle discoverable is not some confused notion of standard or representation, but a physical state that will cause the manifestations of understanding.

Now, the notion of cause may seem inapposite at this juncture, that is, as that which will first give sense to how the understanding determines the range of manifestations of understanding. None of

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

its ordinary (scientific) surroundings is in place. Wittgenstein does not seem to take the proposal seriously. His interlocutor at §195 says, "I don't mean that what I do now (in grasping a sense) determines the future use *causally* and as a matter of experience, but that in a *queer* way, the use itself is in some sense present," and Wittgenstein agrees that causal links are not at issue.

Nonetheless, one might well want to examine more closely what the imagined causal relations are. For, any relation between understanding and manifestation has little of the invariability that a straightforwardly causal one would entail. I do not produce manifestations of my understanding at all times: I may not want to, I may be bored, I may make a mistake, I may have misheard what you said. Thus, any causal story has to bring in other factors: my "belief that you asked thus-and-so," my "desire to manifest my understanding correctly," and so on. Each of these other factors must be imagined to be scientifically discovered as distinct, definite particular states, presumably independently of the state of understanding. Clearly we have little clue as to how such states will be identified. This shows how much is glossed over in the retort; how weak our grasp is of what we are meant to be imagining.[4]

The second strand in Wittgenstein's efforts to undermine the notion of a particular state of understanding concerns the range of practices that are relevant to ascriptions of understanding. I have been talking, simplistically, of "manifestations of understanding," as if all that were involved were correct uses of the understood word. In fact, a major part of Wittgenstein's considerations are to get us to see the enormous variety of features that play a role in our ascriptions of understanding or lack of understanding, and the dependence of these on particularities of the context in which the ascription is made. There is no short repertoire of behavioral manifestations; there is enormous play in what understanding comes to or not in different circumstances; an ascription of understanding can rely on features of the person, far beyond the immediate manifestations.

[4] Here, in talking about the causal efficacy of presumed understanding states, we are in the neighborhood of Saul Kripke's arguments in *Wittgenstein on Rules and Private Language* (Cambridge: Harvard, 1982). Kripke emphasizes the infinitude of potential responses and the finitude of actual ones, given human limitations. My point here is that the problem is not simply a matter of not giving all the responses. Rather, the giving of any response can be thought of as a causal upshot of a state at all only once a vast variety of other putative states is brought into play, involving many areas of the person's psychic economy.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(Different performances are required of new learners and old hands; character traits, mental states—intentional, emotional, and perceptual—and personal history of the subject can be relevant, etc.) Special circumstances of considerable intricacy can affect an ascription (one might reflect, for example, on the range of ways in which mistakes can affect ascriptions). In short, the differences that can emerge in circumstances are overwhelmingly complex; different cases of understanding may have little in common.

Wittgenstein exploits considerations along these lines in many ways, in his general attack. The variegation and lack of uniformity make clear how far any unitary state is from what we observe in the cases of understanding, and thus emphasizes the leap made when matters are framed in terms of such a state. Any of the features that play a role in an ascription of understanding would have to be construed as evidence for this state; but our grasp on the state-evidence model is cast into doubt by the variability of what and when any particular feature functions as evidence. For we lack any conception of how the state is connected to just those particular things which matter in a particular context. If we say that, in each particular case, there is some particular connection, then all we are saying is that the particular circumstances justify our ascription; the one state drops out. On the other hand, what conception do we have of a general connection that issues in such variability? It looks as if each instance of the state must include, in some sense of 'include', something answering to any feature which might arise in any circumstance, in any other instance. This is a notion of internal standard gone wild. It also looks as if the state has to be able to take account of all possible circumstances; but that is not a well-defined totality.

These are ways, more and less subtle, in which the range and variegation of our practices might be exploited, at a stage in the argument when it is still open what the relation is of the putative state to the manifestations; when we are still talking of the "as yet uncomprehended process in the yet unexplored medium" (§308). With respect to the scientific objection, with its envisaged brain states and its recourse to physical causation as linking state and manifestation, the upshot of that range and variegation is more straightforward. It becomes hard to see how we could even begin to identify the tasks the supposed brain state was meant to do, that is, to specify what causal role qualifies a state to be the state of understanding. The objector claims that we could in principle discover that which causes us to behave in ways that a person who understands does; but now it

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

appears that the latter is simply not definite enough. There is no surveyable checklist of such behaviors, inclusion of any particular behavior depends on a not circumscribable range of features of the situation and on the person's particular circumstances and history. Again, what we are imagining, in talking of a brain state identified as the understanding by its causal location, becomes entirely unclear.

A somewhat different cast can be put on the consideration of our practices as follows. The things that we count as behavior associated with understanding are not grouped together by any scientific unity; we have no grasp on the collection in scientific terms. When we talk about manifestations of understanding, we are seeing, or conceptualizing, the behavior through the concept of understanding; it is only that concept that links them together. Wittgenstein seems to be exploiting this in *Zettel,*[5] §§608–611. The point may be summarized thus. If we look at all of the things that we call manifestations of understanding, we find no physical or physiological regularity in them. Hence, there is no reason to expect a physical systematicity in understanding. If you eliminate all the mental notions, what are you left with by way of the data? What you have left, to do the science on, is very little indeed. (The note Wittgenstein is here sounding has a Quinean ring. In fact, I do see a confluence between the two on this point.)

The point emphasizes how the inclination to think that science might discover an appropriate state rests on an accession to a picture of the mental apparatus, of definite states and processes of some sort; given those states and processes, questions can then be framed of whether they can be tied down physically. If, on the contrary, the only apparatus assumed is physical, then upon scrutiny it becomes clear how little conception we have of what is supposed to be discovered.

It may be objected here, though, that—without acceding to any overblown ideas of the mental—surely one can imagine the discovery of a physical state that a person is in if and only if that person, say, understands the word 'manatee'; a state that is merely coextensive with understanding, with no further structure presumed. This is far less clear than it might seem, however. For what Wittgenstein em-

[5] G. E. M. Anscombe and G. H. von Wright, eds., and G. E. M. Anscombe, trans. (New York: Blackwell, 1967). The same passages appear as §§903–906 of *Remarks on the Philosophy of Psychology, Vol. 1,* G. E. M. Anscombe and G. H. von Wright, eds., and G. E. M. Anscombe, trans. (Chicago: University Press, 1980).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

phasizes—the nonuniformity of our criteria for understanding and their intertwining with much else in our physical, social, and mental lives—represents differences from the models we have of states and processes in physical science. The features we rely upon in treating, e.g., physical substances, features that lead us to call them natural kinds and allow us to generalize from one sample to another, are lacking. Thus, it becomes unclear how anything we discover in some cases in which we would ascribe an understanding of 'manatee' could be justifiably generalized to all cases. There are, in a sense, too many singularities.

Suppose, though, that we can carve out some class of cases which is tractable enough from the standpoint of physical science to support inductions, yet broad enough to be interesting; and suppose further that we discover a physical state common to these cases. Is this a discovery of a state of understanding? For example, suppose we found that, in some cases, an understanding of the word 'manatee' goes along with a disposition to blink upon hearing the word. Now, blinking is unconnected to what is centrally involved in understanding, e.g., correct usage and ability to explain the word; hence, the discovered disposition seems extrinsic to understanding, a mere accidental concomitant, rather than anything we would be inclined to call the state of understanding.

Thus, the connections between the imagined physical state and our criteria for ascribing understanding are crucial. More than a bare coextensiveness claim is needed. But then the considerations elicited by Wittgenstein's investigation all serve to show that anything isolable enough and general enough to be usable for the identification of a particular brain state could reflect little of the range of those criteria. Consequently, it can be argued, any state that is so identified would have a tenuous claim to being the state of understanding which the scientific objector claimed to envisage.

WARREN GOLDFARB

Harvard University

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

WITTGENSTEIN AND THE INNER WORLD*

Ludwig Wittgenstein's later philosophy of mind is often read as hostile to the inner; but that is quite wrong, as, e.g., *Philosophical Investigations*[1] §423 makes clear. I shall suggest that our symposiasts leave too much scope for this misreading, at least in respect of the intentional aspects of the mental: Warren Goldfarb in taking Wittgenstein to deny that understanding a word (say) is a "particular or definite state," and Crispin Wright in equipping Wittgenstein with a position according to which intentional states are—to put it crudely—less robustly *there* than nonintentional mental items.

On Wright's reading, Wittgenstein's purpose, in the case of understanding (and this goes for intentional states generally), is to separate it from the occurrent phenomena of consciousness, on jointly phenomenological and a priori grounds. I shall argue that this reading fails to accommodate passages in which Wittgenstein raises difficulties parallel to the one that shapes his treatment of understanding (notably §386); and that, as Wright understands it, Wittgenstein's a priori argument would defeat its supposed purpose, by threatening the existence of any phenomena of consciousness. The distinction between what is in consciousness and what is not cannot have the thematic significance for Wittgenstein which Wright gives it.

The trouble is with the way Wright orchestrates Wittgenstein's moves. There are two obvious targets, Platonism about meaning and the misconception of sensations attacked in the "private language" dialectic. Wright has these as more or less independent temptations to which our thought is prone—although the first, in the shape of a Platonistic conception of intentions, naturally figures in defense of the second, when it is pressed to turn itself into a theory. But I do not believe we can understand Wittgenstein properly unless we see these two target misconceptions as issuing from a common source. I shall suggest that Wittgenstein's fundamental target is a conception of the inner as a lived refutation of "idealism"—a region where we come face to face with the given, that which is set over against conceptual schemes. This makes the grippingness of the misconception of sensations more intelligible than Wright can, and yields a satisfying

* Abstract of a paper to be presented in an APA symposium on the Thought of Wittgenstein, December 30, commenting on papers by Crispin Wright and Warren Goldfarb; see this JOURNAL, this issue, 622–634 and 635–642, respectively.

[1] G. E. M. Anscombe, trans. (New York: Macmillan, 1953).

0022-362X/89/8611/643/4 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

understanding of Platonism and of Wittgenstein's moves against it. It also makes available a satisfying positive picture of the inner world, without Wright's asymmetry between the intentional and the nonintentional; indeed, Wright's conception of the nonintentional aspects of mental life seems to reflect a version of the very thought that, in my view, Wittgenstein is fundamentally concerned to oppose.

Goldfarb discusses material in which Wittgenstein shows that understanding a word, say, cannot be a definite state, on a certain conception of what a definite state would be, namely, the kind of state with which science can deal. ('Cannot' fits here because science cannot do anything with the idea of internal relations.) One of the manifestations of Platonistic self-deception is missing the force of that 'cannot', and taking what is actually a hopeless conceptual bind for a merely empirical mystery ("the as yet uncomprehended process in the as yet unexplored medium": §308). I have two main objections to Goldfarb's treatment of this material. First, he underrates the extent to which "central conceptual features of our notion of understanding" are in place ahead of Wittgenstein's piecemeal investigations, fueling the misconceptions under attack and accounting for their philosophical resonance. Second, he does not consider something that, given my description of the material, is an obvious next move, namely, "So much the worse for that conception of what a definite state would have to be." A lesson of Wittgenstein's philosophy of mind is that we can dislodge philosophical misconceptions, and reclaim the inner world, populated by definite states and processes, for unphilosophical common sense.

JOHN McDOWELL

University of Pittsburgh

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE IMPRACTICALITY OF IMPARTIALITY*

MORAL and political theorists have shared with many ordinary folk a presumption that our everyday moral thinking is often biased. It is a familiar view that bias distorts moral thinking, and obscures the search for sound moral judgment. An unbiased standpoint is widely regarded as the privileged standpoint for critically reflecting[1] on normative matters ranging from basic principles of morality, politics, even rationality itself, to particular and detailed moral situations. This paper does not challenge that assumption.

Eliminating bias from at least certain levels of moral thinking has long been a goal of moral theory. Unfortunately, the unbiased standpoint has proved to be as elusive as it is alluring. The thesis of this paper is that moral and political theorists who propose methods for achieving an unbiased, or *impartial,* standpoint, as I shall refer to it, typically provide no independent criteria for recognizing whether or not impartiality has really been achieved by the specified methods. One is left with no way to confirm, in practice, that moral thought is genuinely impartial, that is, genuinely free of bias.

My discussion will begin in part I by sketching the elusive impartiality that still challenges moral theory. Part II briefly outlines two contemporary theoretical models of critical moral reflection which have been proposed as yielding impartial normative thought: John Rawls's contractual model and R. M. Hare's method of universalization. In part III, I shall argue that these methods of normative thinking represent psychological and epistemic feats, the achievement of which we have no independent way to confirm.

Part IV discusses an alternative approach to impartiality which involves an analysis of impartial reasons in terms of requirements of universality and neutrality. I argue that this analysis also fails to capture what is involved in the notion of full impartiality in moral thinking.

Rather than try to save either the modeling of an impartial standpoint or the analysis of impartial reasons, I adopt yet a different

* To be presented in an APA symposium on Impartiality, December 30. Howard McGary will comment; see this JOURNAL, this issue, 657/8.

This paper was written with the support of a National Endowment for the Humanities Research Fellowship. I am grateful to Larry May for comments on an earlier draft.

[1] My notion of critical moral thinking is similar to that of R. M. Hare in *Moral Thinking* (New York: Oxford, 1981) [hereafter MT].

0022-362X/89/8611/645–656 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

approach. In part V, I recommend that our definitive notion should not be an abstract (and practically unattainable) ideal of impartiality, but, rather, concrete notions of specific *partialities,* that is, biases whose manifestations in normative thinking can be specifically identified and corrected. Coupled with the empirical assumption that people in general lack privileged access to their own biases, this approach invites us to appreciate the practical intersubjectivity of the enterprise of critical moral thinking, and the crucial role played in it by moral dialogue.

I. DELINEATION OF THE PROBLEM

First, some terminological clarification. A variety of terms have been used somewhat interchangeably to identify the unbiased normative standpoint which is privileged by moral theory. 'Impartial', the term used in this paper, is most common,[2] but 'objective'[3] and 'impersonal'[4] are also used. The term 'objective' has a history of usage, according to which it refers to what is external to mind, consciousness, or subjective awareness, notions which are not pertinent to my discussion. 'Impersonal' is a less ambiguous term than 'objective', but still not fully satisfactory. One might say that an impersonal standpoint is the standpoint of no one in particular. The notion of such a standpoint is not equivalent in meaning to the notion of an unbiased standpoint.

'Impartiality' is the term that most typically means an absence of bias or prejudice.[5] Bias is the unduly favoring of one party, person or group of persons when disputes arise or interests compete. 'Prejudice' can also be used to refer to bias, although it, more carefully, means prejudgment, a judgment formed prior to due consideration. These two notions will be combined in this paper, in which it will be assumed that one goal of moral theory is to determine how to rid

[2] This term is used by, among others: Rawls, *A Theory of Justice* (Cambridge: Harvard, 1971) [hereafter TJ], e.g., pp. 187–190; Hare, MT, e.g., p. 129; and Stephen Darwall, *Impartial Reason* (Ithaca: Cornell, 1984), e.g., ch. 15 and *passim.*

[3] This term is used by, among others: David A. J. Richards, *A Theory of Reasons for Action* (New York: Oxford, 1971), e.g., pp. 84–88; and Thomas Nagel, *The View from Nowhere* (New York: Oxford, 1986), e.g., pp. 3–9. It should be noted that Nagel leaves a significant place in moral thinking for reasons and values that can only be apprehended subjectively, thereby eluding the objective standpoint.

[4] This term is used by, among others: Darwall, e.g., pp. 130–145.

[5] Contra theorists such as Roderick Firth ["Ethical Absolutism and the Ideal Observer," *Philosophy and Phenomenological Research,* XII, 3 (March 1952), p. 340], the impartial standpoint need not be defined as dispassionate or unemotional in order to constitute an unbiased standpoint. Whether or not passion or emotion bias normative thinking is a contingent matter in any given case.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

critical moral thinking of any favoritism that is not based on due consideration, or supported by good reasons.[6]

In moral theory, the notion of impartiality itself admits of at least two distinct applications, only one of which is of concern here. Besides defining the privileged standpoint that we are supposed to adopt when critically reflecting on moral matters, impartiality also identifies a duty we are sometimes to incorporate into our daily activities and practices. In this other application, impartiality is a specific duty, for example, of those who fill roles of public trust which involve the dispensing of goods among competing applicants. As a specific duty, impartiality is not always morally obligatory. Some tendencies to favor certain others, such as family members and friends, in one's day-to-day care, assistance, attentiveness, loyalty, and so on, may well receive the imprimatur of unbiased critical moral reflection.[7] But this issue will not concern us in here. My focus is on impartiality as a privileged standpoint for critical moral thinking, and not on impartiality as a specific moral duty.

II. "IF I WERE YOU . . ."

The various contemporary theoretical models of impartial moral thinking include methods of universalization and models of social contract. The requirement of universalizability is well-exemplified in the work of Hare.[8] In Hare's view, moral prescriptions and moral principles have to be universal; it is required by the logical properties of moral concepts (MT, 191). To determine whether a proposed principle or prescription has this feature, one attempts to universalize it: that is, one attempts to determine if one would make the same judgment about any other situation, real or hypothetical, that was identical in its "universal descriptive properties" to the one under consideration (MT, 21).

[6] This paper leaves the core notions of "due consideration" and "good reasons" undeveloped, but acknowledges the risk of circularity in their further specification, a point recognized by philosophy department members at Kansas State University when I read this paper there on April 19, 1989.

[7] Many contemporary theorists have defended the view that impartiality is not morally required in all human behavior; cf. Lawrence Blum, *Friendship, Altruism and Morality* (Boston: Routledge & Kegan Paul, 1980), pp. 46–57. Hare has also defended this view, MT, pp. 135–140.

[8] Hare has developed his views in numerous writings, including his earlier books, *The Language of Morals* (New York: Oxford, 1952) and *Freedom and Reason* (New York: Oxford, 1963). My discussion concentrates on the version presented in his latest book, MT (see fn. 1). The selective focus of this paper involves bypassing the approaches to universalizing presented by other philosophers. Among contemporary theorists, these include: Marcus Singer, *Generalization in Ethics* (New York: Knopf, 1961); and Alan Gewirth, *Reason and Morality* (Chicago: University Press, 1978).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Since the range of all cases identical to this one includes those situations, real or hypothetical, in which one would occupy each of the different positions involved in this situation, one needs to imagine what it is like to be affected as would *each person* in this situation by the action under consideration. One is required to understand what the situation means from each distinct standpoint of identity, preferences, and so on (MT, 91–95). One is to project oneself imaginatively into the perspectives of others, and to attempt to comprehend the situation from their points of view.[9]

The contractual model operates rather differently from the method of universalization. It begins by requiring unanimous consent among the members of the relevant community to a set of basic principles for arranging fundamental moral or social institutions. According to the well-known Rawlsian contractual model for determining basic principles of justice, normative reasoners are: first, to suppress knowledge of their own subjective particulars; second, to suppose themselves possessed of all necessary general knowledge about persons and society; and, third, to abide by certain motivational constraints, most notably, mutual disinterest and lack of envy (TJ pp. 136–142, 538–541). The choices sought from this original position must be ones to which everyone, reasoning under the same constraints, would consent. Unanimous agreement is the sign that no partial interests have been favored by the conclusions.

Rawls's contractual model and Hare's method of universalization contrast in noteworthy ways.[10] The contractual model requires one to diminish one's motivations by eliminating concern for others; by contrast, universalization requires one to comprehend the motivations of others so that one becomes equally concerned with the interests of others as with one's own interests. The contractual model requires one to ignore all particular knowledge about oneself; universalization, on the other hand, requires one to acquire and take account of all relevant particular knowledge about all other involved participants. Both methods suppose that one's relevant general knowledge is comprehensive. And both methods share one theoretically crucial aim: in their different modifications of motive and

[9] Lawrence Kohlberg termed this method for achieving impartiality 'moral musical chairs'; "The Claim to Adequacy of a Highest Stage of Moral Judgment," this JOURNAL, LXX, 18 (October 25, 1973): 630–646; here, p. 641; reprinted as "Justice as Reversibility: The Claim to Moral Adequacy of a Highest State of Moral Judgment," in *The Philosophy of Moral Development* (San Francisco: Harper & Row, 1981), pp. 190–226; here, p. 199.

[10] The subsequent comparison owes a debt to a similar discussion by Richards, pp. 87/8.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

knowledge, each aims to enable one person's thinking to represent the theoretically relevant interests of all involved persons.

The method of universalization presupposes that impartiality results from the incorporation of all perspectives into one standpoint. That is, the universalizing reasoner must somehow compile the views of all standpoints in reaching her normative conclusions. By contrast, the contracting reasoner achieves impartiality by denying the self any self-identifying information that might be used to privilege itself in its calculations. The contractual reasoner is a "stripped down" person, deprived of any individuating self-knowledge, who thereby represents all particular involved standpoints by modeling only what they share in common.[11]

Both of these methods exemplify what Jürgen Habermas[12] has called a "monological" approach to the privileged normative standpoint. They describe a discipline of thought which may be practiced by an individual moral reasoner in isolation. According to the models of both universalization and contract, intersubjective communication is not theoretically recognized to contribute in any special way to the quest for impartiality.[13]

III. THE IMPRACTICALITY OF IMPARTIALITY

The methods of moral thinking required by both the contractual model and universalizability constitute excessively demanding psychological/cognitive feats. My primary objection to these methods is that we can not even be reasonably sure in practice that we have met such requirements. Neither of these theoretical approaches provides

[11] Susan Moller Okin argues that the original position is best interpreted as requiring us each to think from, and care about, the position of "*everybody*, in the sense of *each in turn*" ["Reason and Feeling in Thinking about Justice," *Ethics*, XCIX, 2 (January 1989), pp. 244–246]. This interpretation likens Rawls's method to that of Hare's. An alternative, and, in my view, more accurate reading of *A Theory of Justice* is that of Seyla Benhabib, who charges the original position with lacking any "real *plurality* of perspectives" ["The Generalized and the Concrete Other: The Kohlberg-Gilligan Controversy and Moral Theory," in Eva Feder Kittay and Diana T. Meyers, eds., *Women and Moral Theory* (Totowa, NJ: Rowman & Littlefield, 1987), p. 166]; at most, Rawls differentiates only the "standpoint of the least advantaged representative man" (p. 151).

[12] Cf. "Moral Development and Ego Identity," in *Communication and the Evolution of Society*, Thomas McCarthy, trans. (Boston: Beacon, 1979), p. 90; and Benhabib, p. 167.

[13] For discussion of alternative "dialogical" approaches to moral reasoning, cf. Habermas; Benhabib; and Iris Marion Young, "Impartiality and the Civic Public: Some Implications of Feminist Critiques of Moral and Political Theory," *Praxis International*, V, 4 (January 1986): 381–401. For reservations about a "dialogical" approach, cf. Bruce Ackerman, "Why Dialogue?," this JOURNAL, LXXXVI, 1 (January 1989): 5–22.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

us with the means to determine whether or not the resulting thinking is genuinely impartial, that is, genuinely free of bias or prejudice.[14]

Obviously, it is not within someone's capacity to alter her isolated thoughts so as to encompass general or particular knowledge that she does not, in fact, possess. This feature of either model is certainly unattainable. Instructions of this sort point toward a valuable habit of daily life, namely, the habit of searching for more knowledge relevant to moral matters. And they may well identify a presupposition held by those moral reasoners who claim impartiality, namely, the presupposition that all relevant knowledge has been canvassed. But they are not, in themselves, theoretical exercises that one could perform "at will." This limitation already diminishes the practical value of our two models, but let us (charitably) ignore that fact. Let us suppose that the requirement of augmented knowledge is not critical to either model. We may then examine what remains without this requirement.

This leaves us with two modified models of exercises which we might be able to implement "at will" in order to achieve impartiality: first, imaginative projection into the standpoints of others plus compilation of the results (modified universalizing); and, second, disregard of particular information about oneself and certain of one's motives so as to reason only from what is shared in common with others (modified contracting). Unfortunately, these modified exercises still demand extraordinary cognitive feats. It is not clear how a real person, motivated in the complex ways of ordinary folks, can be sure that she really has either adopted someone else's motivational orientation or disregarded some of her own motivations. Nor is it clear how a person can be sure that she has disregarded her own self-identifying particular information.

Let us examine the requirement to consider matters from someone else's point of view, the lynchpin of Hare's method of universalization. This requirement calls, in effect, for empathy with others. Under everyday circumstances, many of us do empathize with others to varying extents. When we do so, it is usually because we have had similar experiences, or learned about them in great detail, or because the person herself is so familiar to us that we can participate vicariously in her reactions, her attitudes. When we lack personal familiarity or detailed understanding of similar experiences, however, then empathy has little foothold.

[14] Whether or not any of those conditions (altered motivation, altered knowledge, or imaginative projection) is *necessary* for the absence of bias is an issue that I shall bypass.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Thus, imagine the plight of a single mother of four who has no income other than welfare benefits. What attitude will she have toward proposed government decreases in welfare support? She might welcome decreased welfare payments as an incentive to find a job, or she might rue them as an unwarranted penalty on herself and her children for economic circumstances beyond their control. In order to know whether or not she will adopt either of these attitudes, or some other attitude, we must know something of what she presupposes about current economic conditions and about individual responsibility within and for those conditions. We must also know her attitudes regarding the entitlement of dependent children to certain minimum material conditions for growth and development, and regarding the role of the state in assuring that those needs are met.

In general, there are at least two practical problems in implementing the method of universalization. First, what we know about the standpoints of most other persons underdescribes those standpoints. The reference points of being on welfare and contemplating decreased support payments, for example, do not sufficiently define a personal standpoint for moral reflection. An underspecified description of someone else's standpoint does not, by itself, allow one to reason from it to determinate conclusions. Without knowing the motivational complex, basic presumptions, conceptual organization, and so on, in terms of which someone regards her circumstances—without knowing someone's particular point of view in substantive detail—it is simply not possible to consider matters *from* her point of view. Most of the people whose standpoints we have to consider for the purposes of normative theory are strangers known to us only under limited circumstantial descriptions that underdescribe the standpoints of the persons involved.

A second practical problem with the requirement to comprehend a situation from someone else's point of view—a problem which is a matter of degree only and not an absolute limitation—is that someone's motives or preferences, even if known, might be very unfamiliar, alien, even despised from the standpoint of the would-be impartial reasoner. It is not clear that one can really think from the standpoint of motives with which one is unsympathetic or whose subjective force one has never felt. Can I really adopt the perspective of someone who is motivated to molest children sexually? Or to sacrifice German shepherds in satanic cult rituals? Or to incinerate Jews in gas ovens? How does one imagine all the subtleties and extrapolations of motives or attitudes that one despises or does not understand? Without a comprehending familiarity, one can hardly think from the standpoint of alien motives and attitudes except in truncated parody.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The extremely limited nature of intersubjective familiarity poses problems for any theory that sets universalizing, by means of empathy with others, as a condition of attaining the appropriate standpoint for critical moral reasoning. Hare discusses this experiential limitation, but considers it to be a practical, and not a theoretical, difficulty.[15] Such an attitude seems to presuppose that practical difficulties in implementing a moral theory have no bearing on the substance of the theory. This is an unduly narrow view of the proper role of moral theory. If no practical guidelines for implementing a moral theory accompany that theory, this should alert us to the possible practical uselessness of the theory. A moral theory that calls for extraordinary—often impossible—feats of imaginative projection into the standpoints of others does not illuminate a practicable notion of impartiality.

The contractual model bypasses the unrealistic requirement of extensive empathy by not requiring imaginative projection into anyone else's standpoint. Does this alternative approach fare any better than Hare's universalizing? For different reasons, I believe that it does not.

The major difficulty with achieving impartiality on the contractual model has to do with the requirements to ignore certain of one's motivations and all particular knowledge about how one would be affected by the moral arrangements under consideration. The problem is that there is no way to ensure that one's normative thinking is not being tacitly affected by one's own subjective particulars or debarred motivations. Stephen Darwall[16] has pointed out that it is not enough to require that there be no *reference* to self-identifying particular information by the reasoners in a Rawlsian-type original position. In addition, such information must not carry any implicit *motive force.*

Darwall has discerned that the veil of "ignorance" must also be what we might call a veil of "*influence,*" in order to achieve genuine impartiality. The problem is that one's particular gender, race, class background, religion, and so forth, can have motivational influence on one's normative thinking, even when those particulars are not specifically referred to by one's reasons or one's conclusions. Beliefs about the causes of poverty and about society-wide responsibilities to

[15] Examples of "theoretical" problems in determining the preferences of others, which Hare discusses, are the problem of other minds and the problem of interpersonal utility comparisons (MT, ch. 5).

[16] *Op. cit.*, p. 231. On this point, Darwall cites Margarita Levin, "The Problem of Knowledge in the Original Position," *Auslegung*, v (1978): 147–160.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the poor may well reflect the motive force of one's own class background. Beliefs about the legitimacy of subsidized daycare at professional meetings may reflect the motive force of one's parental status. One's degree of tolerance for sexual and other sorts of social violence may reflect the motive force of one's gender.

More recently, Rawls himself has reinterpreted the original position and veil of ignorance as a device for representing legitimate public dialogue in a society with traditions of constitutional democracy. Thus, Rawls now avowedly restricts the application of his theory of justice to people whose particulars of historical tradition precommit them to allegiance to a certain conception of society, namely, to "society as a system of fair social cooperation between free and equal persons."[17] Gone is the argument that choice behind the veil of ignorance constitutes rational choice by a self-interested chooser whose identifying particulars are entirely unknown to her (*ibid.*, p. 237). Her choice is now characterized by Rawls as historically and culturally specific. In that regard, choice behind the veil of ignorance is biased toward the ideals of a particular political culture; to that extent, it is no longer the model of genuine impartiality.

Although Rawls no longer justifies his principles of justice by appeal to the reasoning of a "stripped down" chooser deprived of *all* self-identifying particular knowledge, this device for representing privileged moral thinking still lives on in the work of other philosophers.[18] If we are to constrain our thinking by ignoring all self-identifying personal particulars, then we need some way to determine whether or not those personal particulars are exerting any tacit motive force on our conclusions. Otherwise, we cannot know whether or not the conclusions we have reached are genuinely impartial.

IV. IMPARTIALITY AS UNIVERSALITY PLUS NEUTRALITY

A different approach to impartiality is taken by theorists who, instead of trying to define impartial methods of thought, present an analysis of what would count as impartial propositions, reasons, preferences, or understanding. A typical contemporary analysis involves two requirements: first, reasons must apply to all persons falling within the

[17] "Justice as Fairness: Political not Metaphysical," *Philosophy and Public Affairs*, XIV, 3 (Summer 1985): 225–231; here, p. 229 and passim.

[18] What I am calling "stripped down" thinking approximates what Nagel calls the "objective standpoint." It is the standpoint of no one in particular, the "view from nowhere." Nagel makes some attempt to explicate how such thinking is achieved. The process is said to involve "stepping outside ourselves," "transcending the appearances," transcending "one's time and place," and escaping "the specific contingencies of one's creaturely point of view" (respectively, pp. 140, 139, 187, 9). Nagel's explications are clearly metaphoric and equally clearly unilluminating.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

relevant domain; and, second, reasons must lack essential reference to particular or specifiable persons. The second condition is of special concern here. Let us call it the *neutrality condition.*

The neutrality condition is formulated in various ways by different theorists.[19] Adrian Piper's[20] formulation of the neutrality condition, which she calls "generality," is especially apt. Her formulation calls for the complete absence of proper names and definite descriptions, so that no one, not the reasoner herself nor any other individual, may be particularly identified by reasons that are to count as impartial (*ibid.,* p. 102).

Certain sorts of biases survive Piper's formulation, however. Of great contemporary social interest are biases toward or against whole groups of persons, biases such as misogyny, white supremacism, and homophobia.[21] To foreclose the possibility of these biases, which transcend partiality toward or against particular individual persons, we need a formulation of the neutrality condition which requires the absence of any reference to social groups.

This reformulation would still face at least one remaining difficulty, however. Even if moral reasons were facially neutral, that is, lacking in any essential reference to any individual persons or social groups, it would still be possible for the acceptance or advocacy of those reasons to manifest covert biases. Reasons can be couched in terms which are universal and neutral, yet still advance special interests. Certain "facts" make this possible, for example, facts about the traits that correlate highly with certain types of persons, facts about pre-existing social inequalities, and facts about the ambiguities of facially neutral terms.

One can reason, for instance, that only tall people or only light-skinned people should occupy certain privileged economic or social positions. The "facts" are such that a minimum height requirement will privilege more men than women, and a skin color requirement will have an analogous kind of significance in the area of race.[22] Or

[19] Cf. Nagel, pp. 152/3, who calls it "agent-neutrality," borrowing Derek Parfit's term for Nagel's own idea [Parfit, *Reasons and Persons* (New York: Oxford, 1984), p. 143]. Cf. also Darwall, p. 133, who calls it "impersonality."

[20] "Moral Theory and Moral Alienation," this JOURNAL, LXXXIV, 2 (February 1987): 102–118; here, p. 102.

[21] Nagel's requirement, specified as "agent-neutrality," would handle this problem better than Piper's, if groups were moral agents. This point is controversial, however. Cf. Larry May, *The Morality of Groups* (Notre Dame: University Press, 1987).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

one can advocate racial segregation in facially unbiased terms. Again, the "facts" of pre-existing and entrenched inequalities of resources and power among different racial groups in our culture would combine with such a policy to favor certain racial interests over others.[22] Or one can use male nouns and pronouns as if they were gender-neutral terms. This sadly persistent linguistic convention permits the users of natural languages to mask covert misogynist or male-oriented biases under the guise of a purported gender neutrality.

Recall Darwall's insight that to be genuinely impartial, reasons and conclusions must not only not refer to subjective particulars; they must also not be influenced by the motive force of those subjective particulars. Even reasons which lacked proper names, definite descriptions, and references to social groups could still manifest covert biases because of certain facts about individuals, groups, or the seemingly neutral language used to refer to them.

Rather than attempt to modify the analysis of impartial reasons to handle this problem, I shall, in the next and final section, briefly propose a different approach to the elimination of bias from critical moral thinking.

V. ELIMINATING BIASES

My own proposal is that the abstract ideal of impartiality should not be the primary conceptual reference point around which to orient methods for eliminating bias from critical moral thinking. Instead, our conceptual reference points should be particular forms of *partiality,* that is, named biases whose distorting effects on moral thinking we recognize, and whose manifestations in moral attitudes and behavior can be specifically identified. Our methods for improving critical moral thinking should then involve whatever is needed to eliminate those particular recognized biases. Impartiality is hereby reconstrued as a matter of degree, attainable only incrementally, through the gradual elimination of identifiable partialities, one by one.

As for the methods for eliminating recognizable biases from critical moral thinking, foremost emphasis must go to interpersonal and public dialogue. For good psychological reasons, each person's unaided thinking cannot be trusted to discern its own biases. One's own

[22] This is the kind of discrimination which, in law, is called "disparate impact." It contrasts with explicitly intentional discrimination, known as "disparate treatment." Cf. Claire Sherman, *Sex Discrimination in a Nutshell* (St. Paul, MN: West, 1982), pp. 220–253.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

thinking—explicit and implicit, avowed and tacit—is not fully transparent to oneself. Biases are recognizable, for example, in the particular moral problems which attract or which *escape* one's attention, by the metaphors chosen to express oneself, and by one's reactions to different sorts of persons. The beneficiaries, victims, observers, and so on, of one's behavior may be better situated than oneself to discern biases hidden behind one's articulated moral attitudes, because they can comprehend those avowals contextually in the light of one's related actions and practices.

Persons who are disfavored by a bias, who have the most to lose by the social implementation of that disfavor, are more likely than other persons to be sensitive to its presence in moral thinking and practice. Such persons must be—must have the resources to be—able to participate in public dialogue with people of competing interests so as to ensure that those competitors do not influence social policy and social arrangements in a biased direction. It is notorious that certain groups in our culture lack the resources and status to influence public dialogue in proportion to their numbers in the population.[23] If (as I have assumed) critical moral thinking should be impartial, if (as I have proposed) we can only approach impartiality through the elimination of particular recognizable biases, and if actual dialogue among real persons is a practically necessary forum for this process, then the following result emerges: the real social circumstances of different people and different social groups who can or cannot participate in moral dialogue become urgent matters for any practicable moral theorizing.

MARILYN FRIEDMAN

Purdue University

[23] For an insightful and programmatic discussion of these issues, cf. Iris Marion Young, "Polity and Group Difference: A Critique of the Ideal of Universal Citizenship," *Ethics*, XCIX, 2 (January 1989): 250–274.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

FRIEDMAN ON IMPARTIALITY AND PRACTICALITY*

Marilyn Friedman argues, in "The Impracticality of Impartiality," that various attempts to find an "impartial" perspective from which to evaluate moral judgments cannot succeed. According to Friedman, these approaches fail because they involve epistemic and psychological accomplishments that we have no reliable way of confirming. It is her contention that any moral theory that is unable to give us real guidance in our practical lives does not suffice. So, her main criticism is not that these approaches cannot produce theoretically consistent models of critical moral reflection, but that such models will be of little assistance in helping us to decide, in fact, in an impartial manner.

Alternative approaches, which involve an analysis of reasons and preferences in terms of requirements of neutrality and universality, face the same shortcoming, according to Friedman. Even if the reasons of individual decision satisfy these constraints, as defined by various authors, biases still survive their formulations. For Friedman, this is so because much of the bias that occurs when individuals make moral judgments results because of bias toward social groups rather than toward individuals. So, she argues that an impartial reason requires the absence of any reference to social groups. After rejecting various approaches to finding an abstract ideal of impartiality, Friedman offers her own approach. Her approach focuses on particular forms of partiality rather than on abstract ideals of impartiality.

Although I share some of Friedman's doubts about finding an impartial perspective or impartial reasons for evaluating moral judgments, I would like to press her for clarification on several of her key points. First, I wish to take issue with her reading of what the impartiality theorists are after in their quest for impartiality. Unlike Friedman, I do not believe that the ideal of impartiality requires those who are impartial to empathize with the positions or points of view of others. In order to empathize with another person or her position, one must be able to share her feelings. This requirement, as a condition of impartiality, is too strong. It is rare when we are able to share the feelings of others, especially strangers. Since many moral judgments require that we make decisions involving persons who are not close to us, it is doubtful that we will be able to empa-

* Abstract of a paper to be presented in an APA Symposium on the Impracticality of Impartiality, December 30, commenting on a paper by Marilyn Friedman, this JOURNAL, this issue, 645–656.

0022-362X/89/8611/657/8 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

thize with a vast number of persons with quite different points of view. If this is so, then it is unlikely that we can be impartial when it comes to evaluating persons and positions that are foreign to us.

Second, I question her move to an approach that focuses on particular forms of partiality. I am aware that this way of looking at impartiality is not novel, however. Breaking down specific biases that distort our moral thinking is a popular method adopted by nonphilosophers who attempt to arrive at impartiality. For example, writers who know that they are biased when it comes to evaluating the work of people who have been critical of their work will often request that they be replaced by a judge who does not have such a history with the person whose work is being evaluated. A major problem with this method, however, is identifying the particular forms of partiality which should be eliminated. Friedman is aware of this problem and attempts to address it, but I shall argue that her reliance on interpersonal and public dialogue in generating a reliable method for eliminating biases from critical moral thinking does not suffice. Given Friedman's method, we can eliminate some forms of recognizable partiality, but she does not provide a means for determining how many biases still remain and the role these remaining biases might play in our critical moral thinking. Friedman needs a criterion for identifying forms of partiality which negatively impact on critical moral thinking. Lacking such a criterion or method, I argue that her approach can also be criticized on grounds of impracticality.

Finally, I argue that Friedman's criticism of the abstract-impartiality model may lead to a more radical critique than Friedman offers. If Friedman is right that the abstract ideal of impartiality advanced by thinkers like R. M. Hare and John Rawls is not confirmably attainable through any of the various techniques offered by a host of impartial theorists, and I am correct in maintaining that her move to a dialectical method for eliminating particular forms of partiality can also be described as impractical, then she may be forced to question the very idea of a knowable objective morality.

HOWARD McGARY

Rutgers University

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

HOW MARXISM BECAME ANALYTIC*

> The fact that philosophy cannot finally totalize the alterity of meaning in some final presence or simultaneity is not for me a deficiency or fault. Or to put it another way, the best thing about philosophy is that it fails.
>
> —Emmanuel Levinas

ALTHOUGH Karl Marx offered many social-theoretical explanations, he never made fully clear the form of explanatory model employed in his works. Marx never fulfilled his promise to show "what is rational in the method which Hegel discovered and at the same time mystified."[1] Later Marxists have often been no more clear on this issue, which is the main reason for welcoming the development of analytical Marxism. One of the central conflicts within this new school of Marxism has been whether to retool Marx in terms of methodological individualism (MI) or methodological collectivism (MC). These frameworks are offered by Jon Elster[2] and G. A. Cohen,[3] respectively, as a means to bridge the gap between social structures and individual beliefs and actions. In both cases, the notion of "rationality" plays a crucial role in bridging the gap. Furthermore, these proposals are geared toward the creation of a fully naturalized Marxism. I shall explain these terms in more detail in what follows.

Without minimizing the strengths of these attempts at a naturalized Marxism, my purpose here will be to argue that this form of Marxism is neither fully realizable nor entirely desirable. The guiding subtext here is my conviction that there is too high a price paid by Marx for avoiding that which cannot be naturalized in Hegel, and that this problem is further amplified by the efforts of the analytical Marxists to further de-Hegelianize Marxism.

There are many ways to pursue the argument with naturalism; in the spirit of Marxist immanent critique, I prefer to work through naturalism to that which cannot be naturalized. The steps in this path will be set out as follows: first, a discussion of naturalism and ratio-

* To be presented in an APA symposium on Analytical Marxism, December 29. Andrew Levine will comment; see this JOURNAL, this issue, 667/8. For their helpful comments and criticisms, I would like to thank William Maker and Rex Martin.

[1] Allen Wood, *Karl Marx* (Boston: Routledge & Kegan Paul, 1981), p. 207.

[2] *Making Sense of Marx* (New York: Cambridge, 1985).

[3] *Karl Marx's Theory of History: A Defense* (Princeton: University Press, 1978).

0022-362X/89/8611/659–666 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

nality in Marx and Marxism; second, an outline and critique of the version of MC defended by Cohen, namely, functionalism; third, an account of MI in Elster's *Making Sense of Marx,* the main alternative to Cohen—I take it that Elster has some convincing advantages over Cohen, at least in his critique of functionalism; for my purposes, the important question is whether there is a way to disagree with MI without turning instead to functionalism or losing the advantages that MI has over functionalism; fourth, an outline of the possibility of a nonfunctionalist form of MC—I shall argue that such a notion is not only useful and more powerful than MI as an explanatory model, but indeed fundamental to Marxism; this argument will depend especially on Donald Davidson's philosophy of language.

I. NATURALIZING MARXISM

Marx never succeeded in fully "demystifying" Hegel because of a deep, possibly unresolvable tension in Marxism. This tension has many manifestations in the work of Marx and other Marxists; here I shall simply give a broad characterization of that tension. On the one hand, Marx wanted to create a scientific theory based in empirical studies. One side of Marx's attitude toward Hegel demonstrates what I would call this "naturalizing impulse": Marx sees the dialectical structure of the world as "a complex empirical fact about the nature of material reality . . . not to be explained or understood by means of a priori speculative principles."[4] On the other hand, like Hegel, Marx seeks some basic principle of all reality. But it would be impossible for this principle to be an empirical fact, unless it could be shown that the world is both limited and fully comprehendible. Marxists have traditionally avoided such claims, with good reason. In settling on a basic principle of reality (a claim about the nature of the whole world), however, an a priori, speculative claim is implicitly asserted. This tendency toward metaphysical speculation, which seems a necessary part of any comprehensive theoretical enterprise, I call the *idealizing impulse.* Hegel, in wanting a comprehensive theory, had no choice but to "mystify" that which he also revealed as "rational." The question is whether Marxism, analytical or otherwise, can avoid this tension.

Almost every new way of doing Marxism claims that it has set aside certain metaphysical residues in Marx and other Marxists, and has set critique firmly on the empirical path. In this scenario, "science" conquers "philosophy." The analytical Marxists make this claim in perhaps the strongest form yet, in effect separating themselves off

[4] Wood, p. 209.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

from the entire historical development of Marxism thus far, as a theoretical enterprise and as a form of social activism.

There are notable exceptions to this naturalistic trend: Walter Benjamin, Ernst Bloch, and other Marxists for whom a speculative and even theological overtone is not such an unpleasant sound. These exceptions are important, because they point toward two very significant possibilities. First, perhaps the tension in Marxism is not resolvable—despite the fact that Marx thought that he had resolved the tension in Marxism between "science" and "philosophy" definitively in favor of "science." This fact has had a profound, though not especially happy, effect on the subsequent development of Marxism and socialism. Second, perhaps this tension is what makes Marxism work as a theoretical enterprise and as a social practice in the first place. If this second possibility is true, two alternative paths are again indicated. The preferable alternative, to my mind, would be the recognition that Marxism necessarily contains and, in fact, lives on the empirical/speculative tension. This proposal does not have as a practical consequence a necessary acquiescence toward some finally or permanently established "limits of knowledge"; it is more a question of the better forms of playing the tension out. These better forms would work in accord with the necessities of radically transformative social practice and critique. The other alternative would be to take up some form of critique which avoids this troubling tension.

This is the alternative pursued by the analytical Marxists in their pursuit of naturalism. The analytical Marxists are indeed trying to clarify Marx's own explanatory procedures. The naturalizing impulse comes from Marx himself—and, in Marx, this impulse was always pitted in an adversarial relationship to the idealizing impulse. The analytical Marxists are simply applying the most current analytical tools to the project of pursuing this naturalizing impulse to its definitive end. For Elster in particular, this means not "clarifying" the dialectic, rather, getting rid of it.[5]

There is a history to this antidialectical form of naturalized Marxism.[6] It would be interesting to pursue the question of the basic model of these antidialectical proposals. In every case, the point is to avoid the tension that Marx inherits from Hegel. This tension has

[5] See Tony Smith, "Roemer on Marx and Exploitation: Shortcomings of a Non-Dialectical Approach," in *Science and Society* (forthcoming).

[6] For example, see Otto Neurath, *Wissenschaftliche Weltauffassung: Sozialismus und Logischer Empirismus* (Frankfurt: Suhrkamp, 1979).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

been presented here in an admittedly "atmospheric" way. It is a very complicated question that, for a more comprehensive explication, requires a working out of the problematic of the "other" in Hegel. Hegel ultimately dealt with the problem of the other by assimilating it to the same, to the perfect identity achieved in the final synthesis of absolute spirit. Marx does not discuss the problem per se, but it inhabits Marxism, in one manifestation as the tension between empiricism and metaphysics (naturalism/idealism, science/philosophy, etc.), in another as the social other, the proletariat. In turning to the latest chapters in the story of naturalized Marxism, then, I am especially interested in making sure that the atmospherics of the other remain in play.

In summary, the larger speculation of which this piece of analysis is a part is that the naturalizing impulse cannot contain the idealizing impulse. Indeed, the former actually gives rise to the latter (even if the latter does not have a purely "independent" existence). Furthermore, the naturalizing impulse ultimately propels the idealizing impulse beyond the sort of reclamation that Hegel was after. What is needed, then, is a post-Hegelian, not a pre-Hegelian Marxism, one in which the problematics of the unassimilable other are finally brought alive in a way that they have not been, with only rare but important exceptions, before.

II. COHEN'S FUNCTIONALIST INTERPRETATION OF MARX

Cohen defends the model of base and superstructure typical of Second International Marxism. Although Cohen adds his own twist to this model, defining three levels rather than the traditional two, the interesting thing for our purposes is the "glue" that binds these levels into a coherent whole.

"Higher" levels are functionally explained by "lower" levels: legal relations are functionally explained by social relations, and social relations are functionally explained by the level of development of the productive forces. That x functionally explains y simply means that x has a propensity to bring about y. Any credible functional explanation, however, has to be more specific than this. Cohen's proposal for making "propensities" more concrete consists in two principles: the "development thesis" and the "rationality thesis." The former simply restates the notion of propensity: "the productive forces tend to develop throughout history."[7] The development thesis

[7] Cohen, p. 30.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

must therefore be bolstered by the rationality thesis, helpfully summarized by Alex Callinicos:[8]

> . . . "rational beings who know how to satisfy compelling wants they have will be disposed to seize and employ the means of satisfaction of these wants": given that men are "somewhat rational", that their situation is one of scarcity, and that they "possess intelligence of a kind and degree which enables them to improve their situation", they will expand the productive forces, "for not to do so would be irrational" (*ibid.*, p. 111).

Callincos argues that this rationality thesis (which he calls "transhistorical") cannot ultimately explain the development of the productive forces. But even if the rationality thesis is correct, there are still three problems with giving it such a central explanatory role. First, the thesis seems to amount to a kind of autonomous principle of rationality—there is too much of an "unmoved mover" ring about it. Second, much more can be said about the development of rationality. In part, Cohen's account of rationality is a defense of the notion, much maligned by Marxists, of "human nature." Cohen wants to establish that there are certain "enduring facts" of human nature (pp. 150–153). Perhaps, but Cohen does not go very far in trying to explain rationality without reference to "enduring facts." (In this way, Cohen ontologizes rationality, one indicator of an ahistorical, nondialectical approach, which is further seen in the unidirectional relationships obtaining between the different levels of social organization.) Finally, given the central role that rationality plays in Cohen's theory, if this notion could be more fully explained, then perhaps we would not need functional explanations.

III. ELSTER: RETHINKING MARXISM IN TERMS OF THE RATIONAL-CHOICE MODEL

In *Making Sense of Marx,* Elster criticizes functionalism, offering instead an argument for what he considers the *moral vision* of Marx based in "the search for microfoundations." These foundations are provided, Elster claims, by a "rational-choice explanation of action," which

> . . . involves showing that the action was rational and was performed because it was rational. That the action is rational means that given the beliefs of the agent, the action was the best way for him to realize his plans or desires (p. 15).

[8] *Marxism and Philosophy* (New York: Oxford, 1985).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

There is a close affinity between this doctrine and methodological individualism, by which Elster means

> . . . the doctrine that all social phenomena—their structure and their change—are in principle explicable in ways that only involve individuals —their properties, their goals, their beliefs and their actions. Methodological individualism thus conceived is a form of reductionism. . . . The rationale for reductionism can briefly be stated as follows. If the goal of a science is to explain by means of laws, there is a need to reduce the time-span between explanans and explanandum—between cause and effect—as much as possible, in order to avoid spurious explanations (p. 5).

Functionalism always has a problem in explaining exactly how the overall movement of social forces comes to motivate the individual agent. For Elster, the gap between *explanans* and *explanandum* in functionalist explanations just is not closed by any causal mechanism. Without the ability to show the causal chain leading from productive forces to social relations to legal relations, it is hard to see how Cohen's functionalism is really an explanation. Elster argues that functional explanation is fundamentally misguided—it starts at the wrong end of the chain, and that is the reason there are large gaps in the chain. Cohen's argument seems like a form of naturalized Hegelianism, while Elster's "from the ground up" approach is more like Humean philosophical psychology.

When Elster considers the question of collective solidarity and action, however (which he does not take up until far into the second half of the book), he becomes "both tentative and somewhat less exegetical" (p. 345). Without firmly grounded notions of solidarity, however, it is hard to see what makes Elster's system of ideas "Marxist." Most Marxists have assumed that solidarity grounds autonomy; collective practice and consciousness propels individual thought and action. Both Cohen and Elster have rightly argued that the mechanism of this ground has been inadequately understood. Elster's argument is that the problem is that solidarity cannot come before autonomy—on the contrary. Thus, most of *Making Sense of Marx* concerns the question of autonomous action, or, in Elster's argot, "rational choice."

This problem of relating autonomy and solidarity has been the basis for many attempts to rethink Marxism in a non-Hegelian way. The focus on autonomy is a somewhat typical way to avoid the empirical/metaphysical tension. Jürgen Habermas, for example, is a very

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

important thinker in this regard. Where Habermas moves in a Kantian direction, Elster is more in the line of Hume. Perhaps it would be most appropriate to raise some questions concerning rationality based in the work of Davidson, who is equally indebted to Kant and Hume. Davidson is important here in a further, more important respect, namely, that Elster's rational-choice theory is somewhat indebted to Davidson's action theory.

IV. WHY LANGUAGE CANNOT BE NATURALIZED

From a Davidsonian point of view, Elster's notion of rationality may be interrogated by a question that has troubled reason at least since Descartes. When an agent is formulating her "plans or desires" in light of her beliefs, in what medium are these formulations and considerations carried out? In his experiment in radical skepticism, the one important thing Descartes never questioned was his ability to carry on a mental discussion—Descartes never questioned language. Elster does not question language either, and it is a very important omission.

Elster reads Davidson's action theory (the debt to which is more explicit in *Explaining Technical Change*[9]) in a one-sided way—not an unusual mistake in the interpretation of Davidson's philosophy. Although Davidson often separates, spacially and analytically, his work in action theory and philosophy of language, these two aspects of his philosophy are inseparable. Davidson's action theory avoids the Cartesian end-run around language by specifying that, in the consideration of beliefs and plans, there are always questions of meaning at the heart of the matter.

In "Rational Animals,"[10] Davidson argues that "rationality is a social trait": "[O]nly communicators have it" (*ibid.*, p. 480). His further point is that rationality is generated in and by the medium of communication, language. The larger argument, which undergirds all of Davidson's work, is that it is a shared notion of truth that allows people to communicate. Carol Rovane[11] draws the conclusion that, for Davidson, the bases of subjectivity and intersubjectivity are conceptually identical. My further conclusion is that, since a community of communicators must preceed, temporally and explanatorily, the ability of any particular communicator to communicate, and there-

[9] New York: Cambridge, 1983.
[10] In Ernest LePore, ed., *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson* (Oxford: Basil Blackwell, 1986).
[11] "The Metaphysics of Interpretation," in LePore, pp. 417–429.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

fore to be rational, then solidarity can rightly be said to underpin autonomy. The claim that agents engage in maximizing behavior must be grounded in an analysis of the mental processes involved in the agent's desire formation. Taking language into account, we are compelled to begin explanations of social action one step before Elster, with a linguistically-based methodological collectivism. This form of MC has two major advantages. First, it is not a form of functional explanation. Second, it is capable of comprehending the advantages of Elster's MI. But there is one very big problem.

If we take language to be productive of rationality, then we must now ask if language, rather than rationality, can be fully explained in a systematic way. The problem is that slippage in language (i.e., linguistic indeterminacy) gives rise to slippage in rationality, which Davidson discusses in terms of the "anomolous monism" thesis. (It may be that this thesis could be an opening to the question of the other.) Regardless of how well mental processes are formally understood, we will never be able to formulate precise psychophysical laws, because of the role that meaning plays in the formation of intentions.

The upshot is that the naturalizing impulse will forever be frustrated on some level. This does not mean that there can be no logic to systematic social theory—on the contrary, there must be a very rigorous logic that respects and takes into account the fundamental tension expressive of alterity. That is, Marxism must become "analytic" in a way somewhat different from that suggested by Elster.

The employment of a Davidsonian critique indicates that there is room for the "anomoly" of the idealizing impulse within analytical Marxism. On the other hand, it is doubtful that all readers would necessarily follow me in my assertion that Davidson's proposals concerning rationality and communication lead toward a linguistically-based methodological collectivism. It can be safely asserted that much more argumentation is needed on this question and that, even if this form of nonfunctionalist MC thesis is sound, this collectivism is only the theoretical precondition for solidarity—it is not, in the Marxist sense, solidarity itself. It is very important, however, that we can show that such solidarity is really possible.

BILL MARTIN
University of Kansas

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

REASON AND EXPLANATION*

ACCORDING to Bill Martin, the Marxian theoretical tradition is divided between naturalists, who view Marxism as a scientific enterprise among others, and speculative philosophers who, like Hegel, seek "the basic principles of all reality" beyond science. Martin proposes to defend the speculative philosophers against the naturalists by means of an "immanent critique" that will "work through naturalism to that which cannot be naturalized." He credits "analytical Marxism" with having finally made an immanent critique of the naturalistic project feasible—thanks to its perspicuous focus on "explanatory models." G. A. Cohen's account of functional explanations in historical materialism is, Martin claims, undone and superceded by Jon Elster's methodological individualism and reliance on rational-choice accounts.[1] Elster's explanatory model, too, is untenable, however. According to Martin, it makes good use of Donald Davidson's theory of action. But, when Davidson's philosophy of language is taken into account, the inadequacy of Elster's rational-choice explanations becomes evident; and we are pushed, Martin thinks, toward a "methodological collectivism" consonant with Hegelian speculative philosophy.

I confess that I am not able to say, with any confidence, what it is that Martin wants finally to endorse, in part because I do not understand why he thinks there are "basic principles of . . . reality" that science cannot come to know. I believe that Martin's position is vulnerable in this regard, but I will not pursue this very grandiose question here. There is much to quarrel with, too, in Martin's account of analytical Marxism and its relation to other strains of the Marxian tradition, and in his depiction of the issues in contention in debates about functional explanation. There are also a number of exegetical issues—his account of Elster's views of autonomy and solidarity, and his construal of Davidson's "anomolous monism," among others—where I think Martin goes seriously wrong. I shall pass over all of this, however, in order to focus on a topic that is not merely exegetical, but still relatively tractable and also central to the

* Abstract of a paper to be read in an APA symposium on Analytical Marxism, December 29, commenting on a paper by Bill Martin; see this JOURNAL, this issue, 659–666.

[1] See G. A. Cohen, *Karl Marx's Theory of History: A Defence* (Princeton: University Press, 1978); and Jon Elster, *Making Sense of Marx* (New York: Cambridge, 1985).

0022-362X/89/8611/667/8 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

principal substantive thesis of Martin's paper: his brief against rational-choice explanations in light of what is claimed to be the irreducibly social character of language.

I think that Elster's view of explanation is in fact vulnerable, but not for the reason Martin adduces. To defend this claim, it will be helpful to reflect briefly on the explanatory role of rational-choice models independently of Martin's Davidsonian concerns. I shall argue, first, that there is much less to the dispute between methodological individualism and methodological holism than Martin, following Elster and others, assumes. These positions are either wildly implausible or else they effectively collapse into one another.[2] Nevertheless, Elster's demand for microfoundational (i.e., psychological) accounts of social phenomena—in contrast to the claim that the best explanations in principle for these phenomena are psychological—is warranted for a number of reasons that I shall try to sketch. I shall then turn to the role of rational-choice models in microfoundational accounts, emphasizing affinities between Elster's position and Weber's account of the role of ideal-type concepts in social science. A view of the role of modeling in social scientific explanation will emerge which renders rational-choice explanations immune from questions that can be raised about the universal applicability of the view of rationality assumed in this explanatory genre. I will suggest, finally, that the critique of rationality which Martin thinks he finds in Davidson's theory of language, even if cogent, is as irrelevant to Elster's explanatory program as are other imaginable challenges to the received view.

Elster's professed methodological individualism leads him to speculate in a priori fashion about what the best explanations of social phenomena actually are. This kind of philosophical speculation is illegitimate insofar as scientific explanations are discovered, not prescribed. In any case, there is already good reason to think Elster's prediction wrong. A more open-minded attitude, so far from pushing in the direction of Hegelian philosophy, only reinforces what Martin would have Marxists move "beyond": the naturalistic side of the Marxian theoretical tradition.

ANDREW LEVINE

University of Wisconsin/Madison

[2] Cf. Andrew Levine, Elliott Sober, and Erik Olin Wright, "Marxism and Methodological Individualism," *New Left Review*, CLXII (March/April 1987): 67–84.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE ROLE OF SELF-INTEREST IN ADAM SMITH'S *WEALTH OF NATIONS**

The Wealth of Nations is a stupendous palace erected upon the granite of self-interest.[1]

ADAM SMITH has been called the father of the Industrial Revolution, the greatest political economist, and even the founder of social science. Yet, interestingly, almost since his death there has developed a caricature of his well-known book, the *Wealth of Nations* (WN),[2] a prevailing interpretation which does not accurately represent the content and spirit of the text. What is this interpretation? In the seventeenth century, Thomas Hobbes argued that, deprived of a societal framework that included law and order, each of us in a "state of nature" would be motivated to seek her own self-preservation in disregard of others. If this egoism continues as a motivating force in society, although it is questionable that Hobbes argued that it did, the issue then becomes: How do we restrain, control, and direct our egoistic passions to avoid harm, protect long-term self-interests, and promote the public good? Smith is read as having promulgated this egoistic picture of human motivation in WN, and as having solved the problem of the dichotomy between the so-called natural passions and the public interests. When human beings are granted what Smith calls the "natural liberty" to pursue their own interests, where "all systems of . . . restraint, therefore, being thus completely taken away" (WN, IV.ix.51), the harmony of these individual pursuits will, unintended by the actors, often produce social and economic good. Smith is then interpreted as having concluded that self-interested, economic actors in competition with each other create a self-constraining system through which the impartial market (the famous invisible hand) functions both to regulate self-interests and to produce economic growth and well-being, such that no one actor or group of actors is allowed to take advantage or to take advantage for very long.[3]

* To be presented in an APA symposium on the Thought of Adam Smith, December 30. Charles L. Griswold, Jr. will comment; see this JOURNAL, this issue, 681/2.

[1] George Stigler, "Smith's Travels on the Ship of State," *History of Political Economy*, III (1971), p. 265.

[2] Adam Smith, *The Wealth of Nations* R. H. Campbell and A. S. Skinner, eds. (New York: Oxford, 1976) [hereafter WN].

[3] For versions of this interpretation of Smith, see, for example, George Shack, "Self-Interest and Social Value," *Journal of Value Inquiry*, XVIII (1984): 123–137; A. O. Hirschman, *The Passions and the Interests* (Princeton: University Press,

0022-362X/89/8611/669–680 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

In this paper, I want to question a piece of this kind of characterization of Smith. I want to chip away at the "granite of self-interest." I shall argue that self-interest in WN has the following characteristics: self-interest is usually equated with the natural desire to better one's condition and to look after one's own welfare. Consistent with Smith's earlier work on moral psychology, the *Theory of Moral Sentiments*,[4] self-interest is, however, not to be identified with selfishness or greed. Rather, Smith recognizes the virtue of prudence in the parsimony of one's economic desires. Secondly, in WN, as in TMS, the social passions have equal weight with the selfish ones. This is because (a) self-interest is both driven and restrained by the desire for approval, and (b) economic self-interest makes sense only in the atmosphere of mutual cooperation, a phenomenon which traces its origins to the social passions. Third, the social virtue of justice as fair play, as well as natural liberty, are two norms of a viable well-functioning economy. Smith did indeed present a solution to the problem of how one controls private passions in the public interest, but his solution is not the one ordinarily attributed to him. We shall see how a proper reading of self-interest in the context of his claim that human beings are motivated both by selfish and social passions affects his seminal notion of the invisible hand and thus his theory of a free market.

To get at a proper interpretation of self-interest in WN, one needs to see what Smith was about in TMS. He begins TMS with this statement:

> How selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortune of others, and render their happiness necessary to him, though he derives nothing from it except the pleasure of seeing it . . . (TMS, I.i.1.1).

At the same time, Smith argues, "[e]very man is, no doubt, by nature, first and principally recommended to his own care" (TMS, II.ii.2.1). He explains what appear to be contradictory interests through his analysis of the social passions and his distinctive development of the notion of sympathy. While recognizing that, in the trivial sense, all one's passions are self-interested since they originate in oneself, human beings "by nature" are motivated by three sorts of passions or natural affections: social passions, such as generosity, compassion,

---

1977); Robert H. Frank, *Passions Within Reason* (New York: Norton, 1988); Amitai Etzioni, *The Moral Dimension* (New York: Free Press, 1988); and Milton Myers, *The Soul of Modern Economic Man* (Chicago: University Press, 1983).

[4] A. L. Macfie and D. D. Raphael, eds. (New York: Oxford, 1976) [hereafter TMS].

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

and esteem directed toward others; unsocial passions, such as hate and envy; and selfish passions, such as grief and joy, that are connected with our own pleasure and pain (TMS, I.ii.3–5). What is interesting in Smith's analysis is his claim that the social and unsocial passions have as much weight as the selfish ones.

Smith implies that interests are derived from, but not identical to, the passions. Self-interest or self-love is derived from the selfish passions, but self-love is not identified with selfishness, because self-love, like the other interests, can be virtuous (the virtue of prudence) or evil (greed or avarice). Our interests in others—including the desire for approval, the desire for what ought to be approved of, and sympathy, which is neither a passion nor an interest—come from the social passions. Again these interests can be virtuous (the virtue of benevolence) or evil (harm to others). In addition, Smith distinguishes an impartial virtue of the social passions and interests, the virtue of justice. The unsocial passions translate into reactive interests against others. Often these unsocial interests are not a result of what another has done to harm me, but rather an adverse reaction to actions that have nothing to do with me. Thus, the unsocial interests are distinct from self-interests (TMS, I.iii and II.i.4/5).

In TMS and in WN, Smith uses the term 'self-interest' either to refer to self-love or selfish interests, or simply to designate those interests with which one is most intimately concerned. This is because, he argues, while one's passions are more or less equally directed to oneself and to others, our interests, both our self-interest and interests in others, and accordingly their virtues and vices, are circumscribed by our habits, by those closest to us (e.g., family, friends, and associates), and by the human limits of rationality and foresight.[5] In WN, Smith makes a further distinction between private interests and public interests. Private interests are interests of individuals or groups of individuals that do not concern themselves with society as a whole or the public good. Public interests are those interests which concern society and every member of society equally.

How do these three sets of passions and their corresponding interests play a role in Smith's political economy? According to TMS, individuals are the best judge of their own welfare. In WN, this translates into the natural liberty to "better one's condition" and to take care of one's own economic interests (WN, IV.v.b.16,43). Despite the natural desire to better one's condition which appears to drive economic interests, however, the social passions play three

[5] See Louis Schneider, "Adam Smith on Human Nature and Social Circumstance," in *Adam Smith and Modern Political Economy*, Gerald P. O'Driscoll, Jr., ed. (Ames: Iowa State UP, 1979), pp. 44–69.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

essential roles. In TMS, one of our social passions is the desire for approval by others. In WN, Smith notices that at least part of our economic motivation is not merely to better our own condition, but also to receive approval or admiration (WN, II.iii.28–30). Part of this desire is also envy of another's good fortune (WN, V.i.b.2). Thus, in WN, the desire for approval appropriates the selfish, and the unsocial, as well as the social passions.

The social passions play a second, more important role in WN. Economic interchanges even between self-interested parties are not purely adversary, according to Smith, because they depend on "mutual and reciprocal" coordination. This leads the economist P. L. Danner[6] to contend that such coordination, in turn, arises from the phenomenon of sympathy, a notion Smith develops in TMS.

There are two problems with Danner's theory, however. First, Smith does not use the term 'sympathy' in WN. Secondly, the term 'sympathy', as he develops it in TMS in his technical use of the term, is not a principle of motivation. Rather, sympathy is an understanding of the sentiments of another, as Smith says, a "fellow-feeling with any passion whatever" (TMS, I.i.1.5)[7] of another person, and only some of those "fellow-feelings" might arouse pity, compassion, or empathy. Sympathy is a form of projective understanding whereby I imaginatively comprehend what another is feeling in a particular context. When I find a sentiment with which I both agree and find appropriate in a particular situation, I thereby approve of that sentiment. Sympathy, then, evaluates motivations, sentiments, passions, and actions, but itself is not motivational.

Mutual coordination and cooperation in WN, though crucial to economic exchanges, as Danner points out, are derived from the social passions rather than from sympathy, since it is the social passions that motivate and prompt our interest in, although not necessarily our fellow-understanding or approval of, others. This is clearly evident in Smith's analysis of the importance of the division of labor. He says,

> This division of labor . . . is the necessary, though very slow and gradual consequence of a certain propensity in human nature . . . to truck, barter, and exchange one thing for another (WN, I.ii.1; see also, I.ii.5).

This propensity prompts us to cooperate with others, and it is also to our mutual advantage to do so. "The propensity . . . to truck,

[6] "Sympathy and Exchangeable Value: Keys to Adam Smith's Social Philosophy," *Review of Social Economy*, XXXIV, 6 (1976): 317–331, esp. p. 323.

[7] See also, T. D. Campbell, *Adam Smith's Science of Morals* (London: Allen & Unwin, 1971), ch. 4.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

barter, and exchange" in this context is neither identified with enlightened self-interest nor with benevolence, but is an expression of social interests in economic affairs. Our natural desire to cooperate motivates us to barter with other individuals, while at the same time appealing to their self-interest in the bartering exchange, and, Smith argues, "the gains of both [parties] are mutual and reciprocal" (WN, III.i.1).

To defend the thesis that self-interest is the dominant motivating force in WN, one usually cites Smith's famous remark:

> It is not from the benevolence of the butcher, the brewer, or the baker, that we expect our dinner, but from their regard to their own interest. We address ourselves, not to their humanity but to their self-love, and never talk to them of our own necessities but of their advantages (WN, I.ii.2).

He deliberately uses the "butcher, the brewer, and the baker" example, however, to illustrate the point that, although our self-interests appear to dominate in the economy, cooperation is both natural and essential. None of these tradespeople acts benevolently, yet each depends on our approval for her self-interest. It is to everyone's mutual advantage to produce good bread, beer, and meat, and these tradespeople will fail economically if they do not take this into account. Moreover, in the small-town atmosphere implied in the example, these tradespeople depend on mutual cooperation to stay in business. It is natural in such an atmosphere to desire the approval of others, for example, for the quality of one's bread, and at the same time such approval is an economic necessity. In many other places in WN, too, Smith notices that economic exchanges require cooperation and coordination (e.g., WN, I.ix.16/7, I.x.c.27, and II.v.2/3). In all these cases, cooperation is both natural and required for the development of one's economic interests.

The question is: Do we participate in a cooperative venture because it is in our self-interest to do so, or does cooperation arise out of our social passions and interests and, as it turns out, satisfy the selfish passions as well? Smith is erroneously read to have argued for the former; in TMS, he probably opted for the latter. In WN, he depicts these two sets of passions and their corresponding interests as deeply interrelated. It is both a natural passion and an advantage to cooperate in economic affairs, and one passion does not derive from the other.

Whereas the social passions and interests are crucial in economic relationships, Smith is clear that benevolence has no role. Indeed, we do not expect benevolence in economic situations, for, he says,

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

"[n]obody but a beggar chuses to depend chiefly upon the benevolence of his fellow-citizens" (WN, I.ii.2). Receiving charity is not an honorable economic expectation, nor should it be, as we shall see. Because benevolence is not ordinarily a consideration in economic activities, one commentator has gone so far as to claim that Smith "gave new dignity to greed and a new sanctification to the predatory impulses. . . ."[8] But, in WN, he clearly distinguishes between acting in one's self-interest and greed, and argues that avarice prevents good economic performance (WN, II.iii.25/6). Thus, whereas benevolence may not play a role in economic exchanges, greed is antithetical to a well-run system, and is to be prevented or discouraged. Moreover, according to Smith, economic interests are, or should be, restrained by parsimony, the virtue of prudence (WN, II.ii.36).

The virtue of prudence in economic affairs recalls Smith's depiction of virtue in TMS. There it is clear that self-interest can be virtuous and that benevolence is not the only or even the most basic virtue. According to Smith, virtues are those qualities of excellence singled out by a society as qualities that ought to be approved of or emulated. In brief, he argues that there are four kinds of virtues: prudence, the virtue of restrained self-interest; two virtues of the social interests, benevolence and justice; and a fourth special sort of virtue (a virtue he adds in the sixth edition of TMS), self-command, the virtue of being in command of the unsocial passions and the other virtues. Whether benevolence is the highest virtue in TMS is a matter of debate among Smith scholars, but he clearly states that it is not benevolence that restrains self-interest, but prudence. Smith, then, rather than fuel the egoism-altruism debate, tries to solve that problem with interrelated but distinct sets of passions and interests, each of which has its virtues and evils. Self-interest is neither equated with selfishness nor is it necessarily an evil, nor is altruism or benevolence the only good. Benevolence is not a necessary condition for virtue, and, conversely, a prudent and just person need not be benevolent in order to be virtuous.

The social passions play a third, and crucial role in Smith's depiction of a political economy. In TMS, he argues that justice is the most essential virtue. It arises from the social passions and is the virtue of impartial social interests. Justice in that respect is the "consciousness of ill-desert" (TMS, II.ii.3.4), what an impartial spectator would deem unfair in social relationships. It is both a negative principle that proscribes harming another and includes the positive notion of fair

[8] Max Lerner, "Preface" to Adam Smith, *The Wealth of Nations*, Edwin Cannan, ed. (New York: Modern Library, 1937), p. ix.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

play (TMS, II.ii.i.9 and II.ii.2.1). In TMS, justice is both an individual personal virtue, and "the main pillar that upholds the whole edifice [of human society]" (TMS, II.ii.3.4). In WN, it is primarily the legal system of jurisprudence which is an essential part of the societal institutional framework.

About the scope of justice, Smith says in the *Lectures of Jurisprudence* that the purpose of justice is "to maintain every individual in his perfect rights."[9] Perfect rights are those rights the abrogation of which one is compelled to redress. Since, as he also argues in LJ, one has perfect duties not to harm another nor interfere with their personal liberty, those duties are universally applicable or enforceable. Benevolence, however, is an imperfect duty, since one is not obliged to be benevolent (see LJ, i.14 and v. 142). Whereas the other virtues are vague, voluntarily ascribed to, and not enforceable by law, according to Smith, justice is the only virtue for which one may use external force to ensure. This is because, he contends, "[s]ociety may subsist, though not in the most comfortable state, without beneficence; but the prevalence of injustice must utterly destroy it" (TMS, II.ii.3.3). So justice is the foundation of the other virtues, since it enforces perfect duties without which society could not carry on. Nowhere, however, does Smith develop a notion of distributive justice. This omission is deliberate, because he finds that only commutative justice applies equally to everyone and therefore can be codified and enforced. This is the reason why benevolence should not play a role in economic exchanges. Fair play in competition can be enforced; benevolence cannot.

Justice plays a central role in WN. It is there that Smith is repeatedly critical of laws, regulations, or individual economic actions that are unfair to some participants or group of participants in an economic exchange (WN, IV.iii.a,c; see also, IV.ix.16/7 and I.x.c.27). The norm of fair play, he argues, should govern both one's conduct in economic exchanges and any regulation of that conduct. But he is hardly an anarchist, nor does he hold that the market alone can regulate private interests. Principles of justice have to be codified into law and enforced, according to Smith, because, although most of us do not deliberately set out to harm one another, few of us internalize the ideal of fair play. Although we cooperate with one another in economic affairs, often this cooperation serves private rather than public interests (see WN, I.viii.12, I.ix.24, and I.x.c.25–27). So laws of justice are essential not only to restrain self-interests,

[9] *Lectures on Jurisprudence*, R. L. Meek, D. D. Raphael, and P. G. Stein, eds. (Indianapolis: Liberty Classics, 1978; 1982), i.9 [hereafter LJ].

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

but also because we are naturally cooperative. Therefore, despite the fact that Smith is a strong proponent of the exercise of natural liberty at least in economic affairs, and despite, or because of, the influence of the social passions, he argues that laws of justice are necessary to prevent harms and insure fair play.

If the social passions play a role equal to that of the selfish passions in WN, such that economic exchanges depend on mutual coordination as well as competition, and if laws of justice provide the proper institutional framework without which such activities would be impossible, such a reading obviously affects how one interprets the rest of WN. I shall discuss one topic where this is crucial: Smith's notion of the invisible hand.

Despite the fact that Smith mentions the invisible hand exactly once in WN (and once in TMS), this notion has been the subject of volumes of commentary, interpretation, and extrapolation. To take one example, a well-known commentator on Smith, A. L. Macfie,[10] argues that, in TMS, the impartial spectator, general (moral) rules, an internal sense of fair play, and the virtue of prudence serve as "controls" for human behavior. According to Macfie, in WN, prudence is replaced by the desire to better one's condition, and there is no mention of the impartial spectator or moral rules. The natural social order is replaced by the "natural harmony of the economic order," but there is no guarantee that economic actors will practice internal self-restraint. Therefore, Macfie contends, the invisible hand replaces the impartial spectator in WN as controller of the economic order; it is "the ultimate governor which controls the self-love of individuals . . . the ultimate natural harmony of individual interests" (*ibid.*, p. 101); or even, according to another commentator, Robert Boyden Lamb,[11] "the universal reign of the impartial spectator, the invisible but rational hand, which rational producers for their self-interest must obey" (*ibid.*, p. 681).

The notion of the "impartial spectator" is an important concept in Smith's moral psychology. He wants to create an explanatory mechanism for the possibility of moral judgments which does not depend on a moral-sense theory nor idealizes moral judgments. Just as we can understand fellow feelings, so too, Smith argues, sympathy, coupled with reason, imagination, and our natural desire for "what ought to be approved of" provide the conditions that enable us

[10] *The Individual in Society* (New York: Oxford, 1967).

[11] "Adam Smith's System: Sympathy not Self-Interest," *Journal of the History of Ideas,* XXXV (1974).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

disinterestedly or impartially to examine and to approve or disapprove a motive, character, or action. The impartial spectator, though disinterested, is not necessarily benevolent, since benevolence is only one of the virtues of moral behavior. The impartial spectator is thus the average person in the act of making moral judgments; *average,* according to Smith, because what ought to be approved of is determined by what the general sympathies of a society agree upon as appropriate.

Examining Macfie's interpretation of the invisible hand in light of Smith's concept of the impartial spectator, as I argued earlier, Smith does not replace the prudent person with the desire to better one's condition in WN, as Macfie argues, since he talks about the prudent person and bettering one's condition both in TMS and WN. Moreover, there are crucial differences between the impartial spectator and the invisible hand. The impartial spectator, however objective, is a prototype of, and functions only as, an embodied human being whose judgments are deliberately and self-consciously made by an individual disengaging herself into the role of the spectator. Macfie is correct in saying that, in WN, Smith has less faith in the ability of economic actors to practice self-restraint. But, in WN, this role of personal impartial judgment making is replaced, not by the invisible hand, but by laws of justice.

Although the invisible hand has been interpreted to be the market as it acts impartially to regulate economic behavior, it is in fact merely a result of economic exchanges and cannot function independently. The invisible hand is not the self-conscious deliberate "ultimate governor" of economic behavior or a creator of economic good. It is an explanatory term to describe how economic behavior in mutually competitive and cooperative exchanges can produce unintended economic well-being (WN, I.x.c., II.v.7/8, and IV.ii.9). It is not impartial in the sense that it is an autonomous "hand" that steps back, grasps the economic situation, and dispassionately decides how economic behavior should be balanced and restrained to create economic good. Economic good may be produced as a result of competition. But there is no intentional process that deliberately makes moral judgments about economic good or regulates self-interests "which rational producers for their own self-interest must obey," because there is no "entity" similar to the personification of the impartial spectator to initiate and regulate this process.

If the invisible hand is merely the outcome of economic behavior in a free market, how the invisible hand works and how well it works depend on a number of factors. Counter to a prevailing interpreta-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tion, the market is neither merely driven by, nor functions only because of, self-interested competition. Conversely, market failure is due to a number of factors, not merely overregulation or imperfect competition.

These factors which shape and drive the invisible hand include the following. According to Smith, first, and most obviously, the free market depends on the condition of natural liberty, the absence of constraints on market activities.

> Every man, *as long as he does not violate the laws of justice,* is left perfectly free to pursue his own interest in his own way . . . (WN, IV.ix.51, my italics).

So, second, natural liberty should function only within a framework of justice which protects citizens from harms and guarantees fair play. Smith implies that we do not accept a framework of a system of justice merely because of the authority of law or the utility of the security of our interests. Justice is the impartial virtue of the social passions. So laws of justice which apply to everyone equally are acceptable even when they are not in my self-interest or in the private interests of a group or association. Our objection to laws, according to Smith, is an objection to those regulations which favor or punish some people or groups of people. Thus, it is unfair laws, not regulation per se, that Smith questions (WN, IV.iii.a,c). Third, perfect liberty must be accompanied with self-restraint, an internal sense of prudence and parsimony on the part of economic actors. Unrestrained, selfish, or greedy economic activities can damage competition and hurt the market (e.g., WN, I.xi.p.10, I.viii.13, IV.ii.21, and IV.vii.c.29).

Fourth, and less obviously, a free market works most efficiently when there is competition between more or less equal parties, where one party does not have undue advantages not available to the other. For Smith, the greatest cause of inequality of advantage is any restrictive policy that deliberately gives privileges to certain kinds of businesses, trades, or professions (WN, I.x.c). But he also believes that the market is most efficient and fair when there is a "level playing field," that is, competition between similarly matched parties. Notice that the invisible hand does not create a level playing field, but rather is a result of competition. Smith's ideal is a climate where "[t]he whole of the advantages and disadvantages of the different employments of labour and stock . . . be either perfectly equal or continually tending to equality" (WN, I.x.a.1). This is not to imply that he favors equality of outcomes. Clearly he does not, nor

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

does he think such equality will be a result of a free-market economy (WN, I.x.c.1 and V.i.b.2). But the market is most efficient and most fair when there is competition between similarly matched parties.

Fifth, as I have argued, a market cannot function without mutual coordination or cooperation. The degree of harmony and the extent of mutual cooperation affects whether or not a market functions and functions well. There is a problem with this fifth claim, however. If economic actors are naturally cooperative as well as competitive, why do we need laws of justice? Why is the mutual spirit of cooperation within the framework of competition not enough to balance self-interests and guide the market? With this sort of point in mind, Smith has been said to be an advocate of a form of "benevolent capitalism" as the ideal.[12] There are a number of problems with this conclusion, however. First, it is not benevolence, but fairness, that is desirable in an ideal economy. A benevolent person can be unjust, and benevolence is not enforceable. Second, cooperation and mutual coordination, by themselves, do not guarantee fair play, because people often act in cooperative collusion with each other. Third, in general, according to Smith, people are short-sighted and do not always act rationally. Parochialism in outlook, one's personal situation, and ingrained habits often lead to our placing our own interests and those of our friends, cohorts, and colleagues ahead of the public interest. Even then, because we are neither perfectly rational nor foresighted, we often neglect our own long-term private interests (WN, I.x.c.27 and I.xi.p.8–10). This is why Smith says, "I have never known much good done by those who affected to trade for the publick good" (WN, IV.ii.9). Few of us are capable of understanding what the public good is or how it can best be served, and we often confuse benevolence or charity with fairness and impartiality. This is also why Smith objects to most government regulations—they are not only likely to give advantages to one group over another, but also public servants who administer such regulations are themselves not always impartial or have the proper disinterested perspective. On the other hand, laws of justice which embody principles of commutative justice protect interests equally. Therefore, laws of justice safeguard public interests and restrain cooperative as well as competitive market ventures.

To conclude, self-interest is not the "granite" of WN. Self-interest is not identified with selfishness or greed, and even economic actors are not merely self-interested utility maximizers. They are naturally

[12] David K. Hart, "A Moral Theory of Capitalism," forthcoming.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

prudent and cooperative, and they both seek and need the approval of others. But might it not be the case that, in our rational moments, our economic preferences are self-interested? Smith's answer would be "yes" and "no." Smith notices that economic actors either alone or in collusion with others often act in a self-interested or a group-interested manner. But such activities by and large create monopolies, interfere with free trade, skew competition, and thus disrupt a free market. So a rational person cannot merely act in her own narrow self-interests without threatening the viability of those interests. But, for Smith, it is not merely in my long-term enlightened self-interest to be parsimonious, cooperative, and just. It is also part of my nature to be so, even though I do not act accordingly all the time. For Smith, then, a rational person is prudent, cooperative, and fair, both naturally and because it is to her advantage to be so.

The invisible hand is both more simple and more complex than Macfie envisions. The free market is not an autonomous overseer of self-interest; at the same time, how it functions and whether it functions well depend on a number of complex variables. The market is not self-governing, nor can it increase economic well-being unless all factors affecting it are operating optimally. The invisible hand, then, is a dependent, not an independent, variable in economic activities. Such a conclusion brings into question the philosophical grounds for defending the thesis that an unregulated market can be, in itself, an efficient and impartial regulator of economic behavior, since how one acts in the market place, as well as the character of the framework of justice in which the market is embodied, create and shape the character of the very market that allegedly regulates economic behavior. This conclusion demands a thoughtful reconsideration of some of the defenses of the notion of a self-regulating free market. At the least, neoclassical economic theory might do well to reread what Smith said about self-interest, the market, and the invisible hand.

PATRICIA H. WERHANE

Loyola University of Chicago

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ADAM SMITH ON VIRTUE AND SELF-INTEREST*

> Regard to our own private happiness and interest, too, appear upon many occasions very laudable principles of action. The habits of oeconomy, industry, discretion, attention, and application of thought, are generally supposed to be cultivated from self-interested motives, and at the same time are apprehended to be very praise-worthy qualities, which deserve the esteem and approbation of every body.[1]

In recent years, political philosophy has seen a revival of debate about the merits of classical liberalism. Critics of liberalism frequently insist that the liberal tradition reduces what was a rich, teleologically oriented "person" (as articulated in medieval or Greek traditions, for example) to the status of the mere "individual," an asocial, rootless, utility-maximizing monad by which modern commercial republics are driven. On this general (and here crudely adumbrated) line of reasoning, "liberal individualist politics"[2] is closely allied with—if not simply an extension of—capitalism. Liberalism's "individual" is above all the self-interested actor driven by the passion for the accumulation of property, and capitalism is that system designed to satisfy with maximum efficiency the *pleonexia* that ostensibly defines the citizenry of the liberal enterprise.

Naturally, critics of liberalism who maintain something like the above take an equally dim view of the intellectual progenitor of liberal capitalism, namely, Adam Smith. Patricia H. Werhane quotes Max Lerner as claiming that Smith's *Wealth of Nations* (WN) "gave new dignity to greed and a new sanctification to the predatory impulses." Smith is thus interpreted as a Hobbesian who has discovered the basics of the economic system that best serves the ends Hobbesian agents seek. Werhane rightly aims to call this view of Smith into question. We are thereby invited to reexamine the relationships between virtue and commerce, the praiseworthy and the useful, the public good and self-interest, and, more distantly, the meaning of Smith's high praise of the Stoics. Given Smith's prominent position in the history of the development of liberalism, her argument suggests conclusions that may be of significant relevance to the contem-

* Abstract of a paper to be presented in an APA symposium on the Thought of Adam Smith, December 30, commenting on a paper by Patricia H. Werhane, this JOURNAL, this issue, 669–680.

[1] Adam Smith, *The Theory of Moral Sentiments,* A. L. Macfie and D. D. Raphael, eds. (Indianapolis: Liberty Press, 1982), VII.ii.3.16 (p. 304); hereafter *TMS*.

[2] The phrase is Alasdair MacIntyre's, in *Whose Justice? Which Rationality?* (Notre Dame: University Press, 1988), p. 339.

0022-362X/89/8611/681/2 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

porary debate about liberalism. In my commentary, I shall attempt to draw out these implications.

Although Werhane does not mention it, the question of the role of "self-interest" in Smith's system has been debated for many years, mostly in the framework of the relationship between Smith's TMS and WN. Nineteenth-century German scholarship on Smith specialized in elaborating "*das Adam Smith-Problem,*" complete with psychological and biographical speculations about Smith's supposed personality changes and mental development, speculations intended to explain the supposed tension between the two works. TMS insists from the very first that human beings are fundamentally "social" creatures, whereas WN supposedly insists that we are fundamentally self-interested agents. But this is a pseudo problem; if there is a tension in Smith's position on the altruism/egoism debate, it recurs within each of his books, not between them. This means that his analysis of "capitalism" (a word Smith does not himself use) unfolds within the framework of a moral theory (a central component of which is a critique of "self-love" and greed).

Although Werhane is entirely justified in taking up the "self-interest" issue, I believe that her account of Smith stumbles at several junctures and omits several important dimensions of Smith's analysis. I am also less persuaded than is Werhane that Smith's wonderfully nuanced analyses of moral psychology and of liberal economic and political orders ultimately succeeds in solving the altruism/egoism problem. Much hinges on the meaning of the verb in Werhane's statement that "in WN the desire for approval appropriates the selfish, and the unsocial, as well as the social passions." I shall develop these doubts by pursuing several questions not posed explicitly in Werhane's paper, among which are these: Is Smith's moral psychology normative, descriptive, or both? Does Smithean Stoicism mesh consistently with liberal capitalism?

CHARLES L. GRISWOLD, JR.

Howard University

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

NOTES AND NEWS

The *Journal of the History of Philosophy* Monograph Series announces a call for papers. The series accommodates serious studies in the history of philosophy which are between article length and standard book size. Ranging from ancient Greek to twentieth-century philosophy, the series included extended studies on individual philosophers, ideas, analyses of texts and controversies, new translations and commentaries on them, and documentary findings about various thinkers and events in the history of philosophy. Manuscripts dealing with ancient through medieval philosophy may be directed to Charles M. Young, Phil. Dept., Claremont Graduate School, Claremont, CA 91711; and those dealing with modern and recent philosophy to Richard A. Watson, Phil. Dept., Washington Univ., St. Louis, MO 63130.

The editors of *Teaching Philosophy* wish to announce the appointment of James Campbell as Book Review Editor. Books for review and notes from interested reviewers may be directed to him at the Phil. Dept., Univ. of Toledo, Toledo, OH 43606.

Theoria et Historia Scientiarium, a new interdisciplinary journal that serves as a forum for international academic exchange among researchers of the natural sciences, social sciences, and humanities, announces a call for papers that cultivate historical and interdisciplinary points of view. The journal's editorial and advisory boards include distinguished scholars from Canada, England, Finland, Greece, Poland, and the U.S.A. Papers may be directed to Wieslaw Mincer, Univ. Library, Biblioteka Uniwersytetu Mikolaja Kopernika, Gagarina 13, 87-100 Torun, Poland.

The Program in Ethics and the Professions at Harvard University invites applications for Fellowships in Professional Ethics. Six Fellowships will be awarded in 1990/91 to outstanding teachers and scholars who wish to develop their competence to address ethical issues in schools of business, government, law, and medicine. Fellows will participate in the seminar of the Program, attend courses in one of the professional schools or in the Graduate School of Arts and Sciences, and conduct their own research on professional ethics. The Fellowship usually extends from September through July. Applicants should hold a postgraduate degree in business, government/public policy, law, or medicine, or a doctorate in philosophy, political theory, or theology. Application deadline is January 5, 1990. Further information may be obtained by writing Jean McVeigh, The Program in Ethic and the Professions, Harvard Univ., 79 J.F.K. Street, Cambridge, MA 02138.

Marquette University is pleased to announce its new Center for Ethics Studies. The Center will focus on both theoretical and applied/professional ethics. Inquiries may be directed to William Starr, Center for Ethics Studies, Academic Support Facility, Rm. 336, Marquette Univ., Milwaukee, WI 53233.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The Bertrand Russell Society announces a call for papers to be presented at its meeting at the Eastern Division of the APA in December 1990. Papers may be on any aspect of Russell's philosophy, and may be submitted, in triplicate and prepared for blind review, to David E. Johnson, Bertrand Russell Society, Sampson Hall, U.S. Naval Academy, Annapolis, MD 21402-5044.

The Department of Philosophy at California State University/Long Beach is pleased to announce that Associate Professor Paul C. L. Tang has assumed the chair, and that Roger Wertheimer has been appointed Associate Professor.

The Department of Philosophy at California State University/Northridge is pleased to announce the appointment of Cindy Stern as Assistant Professor.

The Department of Philosophy and Religion at Clemson University is pleased to announce the appointment of Stuart Silvers as Professor and Head of the department.

The Department of Philosophy at Colorado State University is pleased to announce the appointment of Michael Losonsky as Assistant Professor.

The Department of Philosophy at Rutgers University is pleased to announce the senior appointments of Colin McGinn and Stephen Stich, and the junior appointments of Connie Rosati and Mary Frances Egan.

The Department of Philosophy at San Diego State University is pleased to announce the appointment of Jack Zupko (Cornell) as Assistant Professor.

The Department of Philosophy of the State University of New York/Binghamton is pleased to announce the appointments of Eileen Way and Eric Dietrich as Assistant Professors in its Program in Philosophy and Computer and System Sciences.

The Department of Philosophy at the University of California/Davis is pleased to announce the appointments of David Copp and Richard Wollheim as Professors of Philosophy.

The Department of Philosophy at the University of California/Los Angeles is pleased to announce the appointments of Kit Fine and David Pears as Professors of Philosophy.

The Department of Philosophy at Washington University (St. Louis) is pleased to announce that Roger F. Gibson succeeds Lucian Krukowski as Chair. The department would also like to announce that the Visiting Philosophers for 1989/90 will be Marc Ereshefsky, Dan Gilman, Tom Oberdan, and Gary Varner.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVI, NO. 12, DECEMBER 1989

THE PRINCE AND THE PHONE BOOTH: REPORTING PUZZLING BELIEFS*

IN Mark Twain's *The Prince and The Pauper*, Tom Canty and Edward Tudor decide to change lives for a day, but fate intervenes, and the exchange goes on for a considerable period of time. The whole story turns on what people believe and do not believe about the two boys, and an intelligent reader, unexposed to recent philosophy of language and mind, could probably describe the key facts of the story with some confidence. Such a reader might explain why Miles Hendon, a penniless nobleman who encounters a boy dressed in rags, does not bow to the Prince, by noting:

(1) Miles Hendon did not believe that he was of royal blood.

And such a reader might ward off the implication that Miles was a fool or ignoramus by noting that Miles shared the dominant conception of Edward Tudor:

(2) Miles Hendon believed that Edward Tudor was of royal blood.

One of our main claims in this paper is that such a reader would be right on both counts. In this we depart from a recent trend to explain the apparent truth of statements like (1) as an illusion generated by pragmatic features of such claims. Accounts of belief reporting given by Jon Barwise and John Perry,[1] Scott Soames,[2] and Nathan Salmon[3]

* This work was supported in part by the System Development Foundation through a grant to the Center for the Study of Language and Information. We would like to thank the Philosophy of Situation Theory group at CSLI; special thanks to David Israel.

[1] *Situations and Attitudes* (Cambridge: MIT, 1983), pp. 253–264.

[2] "Substitutivity," in *Essays in Honor of Richard Cartwright* (Cambridge: MIT, 1987), pp. 99–132, and "Direct Reference and Propositional Attitudes," in Almog, Perry, Wettstein, eds., *Themes from Kaplan* (New York: Oxford, 1989).

[3] *Frege's Puzzle* (Cambridge: MIT 1986).

0022-362X/89/8612/685–711 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

have employed this strategy of denying the accuracy of our strong intuitions about truth and falsity. Here, we shall present an account that does not ignore pragmatic features, but assigns to them a more honorable role. They do not create an illusion, but help to identify the reality the report is about. Our account honors the intuition that claims (1) and (2) are true.

Since 'Edward Tudor' in (2) and 'he' in (1) both refer to Edward Tudor, this seems to commit us to some version of the doctrine of *opacity*.[4] Specifically, we are committed to the view that, if our reader were to say either of the following, in the same circumstances, he would be incorrect:

(1′) Miles Hendon did not believe that Edward Tudor was of royal blood.

(2′) Miles Hendon believed that he was of royal blood.

The doctrine of opacity has been thought incompatible with two others, to which we also are attracted: the first, *direct reference*, is that the utterance of a simple sentence containing names or demonstratives normally expresses a "singular proposition"—a proposition that contains as constituents the individuals referred to, and not any descriptions of or conditions on them; the second, *semantic innocence*, is that the utterances of the embedded sentences in belief reports express just the propositions they would if not embedded, and these propositions are the contents of the ascribed beliefs.[5]

Direct reference and semantic innocence are well-motivated by many considerations in the philosophy of language. But if direct reference and semantic innocence are correct, then it seems that opacity must not be: the substitution of 'Edward Tudor' for 'he' in (1) [or vice-versa in (2)] should be completely legitimate. The name and the demonstrative refer to the same object. There is just one proposition, belief in which is denied by (1) and affirmed by (2), the "singular" proposition, which we shall represent in this way:

≪Being of royal blood; Edward Tudor≫

[4] *Opacity* is the claim that substitution of coreferring names and demonstratives in belief reports does not necessarily preserve the truth of those reports (definite descriptions are another matter; it is not nearly as controversial that substituting a description for a coreferring name can influence the truth value of a belief report). What "substitution" comes to with respect to utterances (belief reports), as opposed to sentences (belief sentences), is not at all obvious. Our simple notions of substitutivity, opacity, and so on are really useful only if sentences are (wrongly) taken as the bearers of truth and content. Here, we shall adopt an informal notion of substitution in belief reports, such that the reports (1) and (1′), as well as (2) and (2′) are related by substitution.

[5] For an important qualification, see fn. 14 below.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The example is typical of many doxastic puzzle cases in the literature—puzzles because they seem to reveal a conflict among the three very plausible doctrines. We hold all three, however.

I.

When we substitute 'Edward Tudor' for 'he', the words change while the proposition expressed by the embedded sentence stays the same. If we think that belief is a relation to propositions and not words, the apparent change in truth value of the whole report seems puzzling. We are likely to focus on the most apparent change, the change in words, as the clue to the mystery.

The most famous doxastic puzzle case, due to Saul Kripke,[6] has nothing to do with substitution, however. Kripke describes a case in which the Frenchman Pierre first hears of London, comes to believe it is pretty, then moves to London, and, not connecting it to the city he has heard about (under the French 'Londres'), comes to believe it is not pretty. He does not change his mind about the city he has heard of, but simply does not connect the "two" cities. We have one sentence

(3) Pierre believes that London is pretty.

that we seem to be able to use, when reflecting on different parts of the story, to say something true and to say something false. The words have not changed. What has?

What changes in this case, and in every other doxastic puzzle case, is what we are talking about. Pierre has two different notions of London, which play very different roles in his beliefs. An assertion of (3) is true if it is about one of them, false if it is about the other. An ordinary doxastic puzzle case uses a change in words to precipitate the change in the subject matter of the utterance. Kripke spells out the details of his case so clearly that our focus gets redirected without a change in the wording of the report. We shall return to these claims about belief reports in the next section.

One of Pierre's beliefs was caused by his acceptance of the stories he heard about London. It has the content that London is pretty, and it leads him to cherish the prospect of someday visiting that city. This belief also causes him to affirm, in French, "Londres est jolie," in discussions about the city he has heard of.

Also, Pierre has a different belief which was caused by his displeasure with his new surroundings, which has the content that London

[6] "A Puzzle about Belief," in A. Margalit, ed., *Meaning and Use* (Dordrecht: Reidel, 1979), pp. 239–283.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

is not pretty and which causes him to affirm, "London is not pretty," in discussions about his home.

It is a commonplace to distinguish these two beliefs. We think it is often not sufficiently appreciated, however, that the beliefs so distinguished are concrete cognitive structures. Focusing on this fact provides the basis for our account of belief and for our solutions to the various doxastic puzzle cases.

These are the key features of our theory of beliefs:

(i) Beliefs are concrete cognitive structures: they are particulars that belong to an agent, come into existence, endure, and go out of existence.
(ii) Beliefs are related to the world and to other cognitive structures and abilities in a way that allows us to classify them by propositional content.

Beliefs, since they are cognitive particulars or "things in the head," are not things that are believed; they are not in any sense the objects of belief. The propositions believed are the objects of belief. An agent believes some proposition in virtue of having a belief with that content. Many agents can believe the same proposition, so propositions are public; they also are abstract. Beliefs are neither public nor abstract; they are concrete particulars that belong to agents just like arms, headaches, and bouts of the flu. A belief comes into existence when an agent forms it; it is not the sort of thing that is around for the agent to adopt. Agents believe the same thing, a proposition p, when each has a belief with p as its content. This is not an analysis of reports of "believing the same thing"—which are not always so simple to unpack—but a clarification of what we mean by objects of belief.

To countenance beliefs as particulars is not to deny that there are interesting systems of abstract objects which might be used to classify them, such as meanings, Fregean senses, intensions, characters, or the like. But in addition to having these abstract features, beliefs, like other concrete particulars, have lots of other features, both intrinsic and relational, many of which can in some cases be relevant to explaining how we talk about beliefs in belief reports. In particular, we often exploit facts about the causes and effects of beliefs, a point to which we shall return.

There are a number of reasons to allow ourselves to speak of particular beliefs, rather than just of a belief relation between a person and an abstract object of some kind. There is, first, the attraction of having entities that can occupy causal roles with respect to perception, reasoning, and action. As Jerry Fodor and others have argued at length, structured concrete particulars or "token" mental

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

entities go a long way toward explaining the roles of belief, desire, and so on in cognition. There is, second, the fact that the most plausible statements of materialist intuitions about the mind are formulated in terms of particular mental entities. And, third, there is the problem that belief puzzles repeatedly have emphasized: it seems that, for any natural way of classifying beliefs with abstract objects, we can find examples in which a single agent, at a single time, is belief-related to one such abstract object twice over. These are cases, we would like to be able to say, in which an agent, at a time, has two beliefs classified by the same sense, meaning, or whatever. Classifying beliefs only with abstract meanings, senses, and so on is like classifying drops of water only with intrinsic properties. Kant argued against Leibniz that intrinsic properties of particulars will not always provide us with sufficient material for their individuation. Kant took it as obvious that there can be two exactly similar drops of water; the puzzle cases make it clear that there can be two beliefs sharing the abstract features which one or another theory of belief claims to be central.[7]

Beliefs, then, are particulars that bear complex causal relations to an agent's perceptions, actions, and other cognitive structures and abilities. The story of the causal properties of beliefs will be closely bound to the story of how and why beliefs can be classified with propositional content. A belief constrains an agent's reasoning and action in a way that is conducive, if the belief's content is true, to the agent's getting what she wants.

The ground-level facts behind belief are simply the facts of agents having beliefs. There is a basic relation $B(a, b, t)$ that holds of an agent, a belief, and a time just in case b is a belief that belongs to the agent a at time t.

Normally, a belief has a propositional content. So there is a partial function Content(b, t) that, for a belief b and time t at which b exists, yields the content of b. The content of a belief will be determined by the "internal" structural properties of the belief plus its real connections to things and circumstances in the world and to the agent's other cognitive structures and abilities.

If an agent a at time t has as an object of belief the proposition p, then there is a belief b such that:

$$B(a, b, t) \;\&\; \text{Content}(b, t) = p.$$

So much is all that is really needed for a theory of belief adequate for a broad explanation of the doxastic puzzle cases, and so we are

[7] For a fuller defense of the particularity of beliefs, see Crimmins, *Talk About Beliefs*, Ph.D. dissertation, Stanford University, 1989.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tempted to stick with just the minimal theory of beliefs given so far. The minimal theory is compatible with a wide class of views about beliefs, about propositions (or contents), and about central issues in theories of representation, practical reasoning, and inference. The crucial features of the semantics we give for belief reports, and the resulting solutions for the troubling cases, are therefore to some degree theory-neutral. But we want to present a slightly more detailed, if still simple-minded, theory of beliefs which satisfies the demands of the minimal theory and which yields a sufficiently rich account of just how the puzzling belief reports work.

Beliefs are structured entities that contain ideas and notions as constituents. Ideas and notions, like beliefs, are on our view concrete cognitive particulars. So there is no such thing as agents having the same idea or notion, but only similar ones. Admittedly, the technical use we make of these terms involves a departure from what we ordinarily say about 'ideas' and 'notions', or at least represents a choice among the many different ordinary uses of these terms. On our use of the terms, there are no notions and ideas that agents do not have, any more than there are headaches that no one has. The difference between notions and ideas is the difference between an agent's "ways of thinking" about individuals versus properties. The properties and things of which ideas and notions are ideas and notions we call their *contents*. We shall explain in a moment how the contents of ideas and notions help determine the contents of beliefs.

What determines the content of an idea or notion? For example, what is it about Miles's notion of the poorly-dressed boy which causes it to be a notion of Edward as opposed to another boy? The crucial fact is that it was Edward with whom Miles was confronted when he formed this notion. Edward played the right part in the causal origin of the notion; the notion was formed in order to keep track of information about Edward—that is what makes him its content. So the content of an idea can depend on its external properties, like facts about its origin. The very same notion might have been a notion of a different person, had someone other than Edward figured in its origin.

There is a close parallel between this view of the contents of ideas and causal views of the semantics of names. A speaker can refer to an individual with a name, it is held, because that individual figured, in the right way, in the speaker's adoption of the name as a tool of reference.[8]

[8] A speaker can adopt a name (like 'John') more than once, to refer to what may be different individuals. Each such adoption creates a type of use to which the speaker may put the name. So a causal analysis of names should look not at names themselves, but at types of uses of names, as the things for which reference is

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The content of an idea is not always fixed once for all by facts about the circumstances of the idea's origin. Some ideas are *context-sensitive,* in that their contents change with changes in the agent's circumstances. The context sensitivity of ideas is analogous to that of demonstratives in language. David Kaplan[9] has proposed that there is associated with each demonstrative a *character,* a function that specifies how the content of a demonstrative depends on the circumstances surrounding its use. The content of a use of the word 'you', for example, is the person being addressed in the circumstances of the utterance. Analogously, an agent a may have an idea I_{addr} of "being the one I am addressing." The property which is the content of this idea changes with changes in circumstances as follows:

> In any circumstances in which a person b is being addressed by a, the content of a's idea I_{addr} is the property of being b.

Undoubtedly, each of us has a "you" idea, the content of which is determined functionally in this way. We do not share ideas, but we have ideas with the same *semantic role.* An idea's semantic role is the function that determines the idea's content based on the agent's circumstances. Semantic roles for ideas are a bit like characters of expressions; some ideas have semantic roles that are context-sensitive, others have semantic roles that are constant functions—their contents do not vary with changes in context.

So there are two ways in which an agent's external circumstances might be relevant to determining the content of an idea. First, the facts surrounding the origin of the idea may fix its content once for all. Second, the idea's semantic role may be sensitive to changes in the agent's circumstances—the content of the idea may vary from occasion to occasion. So an idea may exhibit origin sensitivity, context sensitivity, or both.

Miles's idea of red is certainly not context-sensitive. It may be deemed origin-sensitive, whether one supposes that his idea stands for red innately, or because of some original assignment of ideas to colors early in Miles's life. Miles's idea of being past, in contrast, stands for different properties as his life unfolds; at each time t, this idea stands for the property of occurring before t. This idea is certainly context sensitive, and may or may not be origin sensitive. And

---

determined causally. An agent may use 'John' to refer either to John Dupre or to John Etchemendy. What individuates these distinct types of uses of the name 'John'? One answer is that the types of uses of 'John' are tied to distinct notions in the agent.

[9] See "Demonstratives," 1977 manuscript reprinted in *Themes from Kaplan.*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Miles's notion of Prince Edward, formed upon hearing of the newborn Prince, is origin-sensitive, but it is not context-sensitive.

Notions are the things in the mind that stand for things in the world. A notion is a part of each of a collection of beliefs[10] (and of other mental structures, such as desires and intentions) that are internally about the same thing. This is not a definition of 'notion', but just a central fact about notions—sharing a notion is what it is for beliefs to be internally about the same thing. An agent may occasionally (and will in many of the examples) have several notions of a single individual. This can happen in two ways. First, in cases of misrecognition and "failure to place," an agent may have two notions of an individual which he does not link or connect; such an agent is guilty of no internal inconsistency. But also an agent can retain two notions of an individual, while linking them, in the way one does when one recognizes that "two" of one's acquaintances are actually a single individual. Why might two notions be retained when such a recognition takes place? One reason for this would be to allow the possibility of easy revision in case the "recognition" was in error. But an agent can also burn his bridges and merge two notions into a single notion. Two beliefs, then, can be internally about the same thing in two ways: by sharing a notion, and by containing notions that are linked.

For the purposes of this paper, we assume that each belief involves a single k-ary idea and a sequence of k notions.[11] To represent the structure of such a belief, we write:

$$\text{Structure}(b) = \langle \text{Idea}^k, \text{Notion}_1, \ldots, \text{Notion}_k \rangle.$$

Each belief has as its content the proposition that the objects its notions are of have the property or stand in the relation, that its idea is of:

$$\text{Content}(b, t) = \ll \text{Of}(\text{Idea}^k, t); \text{Of}(\text{Notion}_1, t), \ldots, \text{Of}(\text{Notion}_k, t) \gg.$$

The structures of beliefs are individuated not simply by the ideas and notions involved in them, but also by which argument places of the ideas the various notions fill. Thus the order of the notions in our representation of the structure of the belief reflects an assignment of notions to the argument places of the associated idea.

[10] There is no mystery as to how a single thing can be a part of many different things at the same time (and at different times). One may, for example, be a member of many different committees or clubs.

[11] This is to consider only beliefs of a certain kind of composition. In a more thorough presentation, a discussion of other kinds of belief structures, perhaps including general beliefs and complex beliefs, might be called for—although the logical connectives and quantifiers can be accommodated within this simple structure. Also, we have chosen to ignore in this paper many subtleties of time and tense.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

To be clear about the relation between beliefs and their contents, we need to introduce some new concepts.

A belief b associates an idea I with a notion n at an argument place pl:

$$\text{Associates}(b, I, n, pl)$$

The belief that Tom fired Mary and the belief that Mary fired Tom differ in which places are associated with which notions, even though the ideas and notions involved are the same.

An argument place of an idea is intimately connected with an argument role of the relation which is the content of the idea, and so with an argument role in the content of the beliefs of which the idea forms a part.[12] If we were to consider complex cases, spelling out this relationship might be a matter of some delicacy, but we shall take it to be straightforward here. We shall say that an argument place pl_I of an idea I generates an argument role r_p of a proposition p (an example below will make this clearer):

$$\text{Generates}(pl_I, r_p)$$

Finally, a notion is responsible for which object occupies an argument role of the content of a belief, when the belief associates it with an idea at the argument place which generates the argument role in the content of the belief:

$$\text{Responsible}(n, r, b) \Leftrightarrow_{def} \exists I, pl \text{ Associates}(b, I, n, pl), \text{ and Generates}(pl, r)$$

When a notion in a belief is responsible for filling an argument role of the belief's content, it fills the role with its own content, the object of which the notion is a notion.

To give an example: Arthur's belief that Yvain smote Kay involves Arthur's idea for smiting, I_s, and his two notions of Yvain and Kay, call these n_Y and n_K. The idea I_s has two argument places, one (pl_+) for the smiter and one (pl_-) for the smitten. In Arthur's belief (call it b), the notion n_Y is associated with argument place pl_+ of I_s, and n_K is associated with pl_-. The content of b is the proposition p, where:

$$p = \ll\text{Smote; Yvain, Kay}\gg$$

The relation "smote" has two argument roles, one (r_+) for the smiter and one (r_-) for the smitten; these are also argument roles of the proposition p. In p, Yvain fills r_+ and Kay fills r_- of the "smote"

[12] Roughly, an argument role of a relation is also an argument role of a proposition (at least) when, in that proposition, the role of the relation is occupied by an object.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

relation. Since b associates pl_+ with n_Y, and pl_+ (the smiter in I_s) generates r_+ (the smiter in p), we say that, in b, n_Y is responsible for filling r_+ in p. Arthur's notion of Yvain is responsible in b for determining who fills the argument role r_+ in p. And n_Y provides its content, Yvain, to fill that argument role. Figure 1 should make this clear.

Notions and ideas are key figures in our commonsense "folk" model of cognition. The recurring appearance in philosophy of such things as concepts, senses, ways of thinking, names in a language of thought, mental file folders, and other such devices reflects a firm intuition about the mind, namely, that having beliefs about an individual means having beliefs involving an internal something that is one's cognitive "fix" on the individual. As we have said, we think the correct way to express this intuition demands reference to cognitive particulars that are involved in beliefs, desires, and so on. Now, this leaves a great deal open about just what kinds of thing our notions and ideas are. For all we have said, notions and ideas might be—or might have been—particular words in a language of thought, physical objects like file folders, or things with more of a dispositional character, like the process underlying the disposition of an agent to have a specific "pattern of neural activation" in certain circumstances. And, whichever of these kinds of thing our notions and ideas are, they certainly may be classifiable with senses, property clusters,

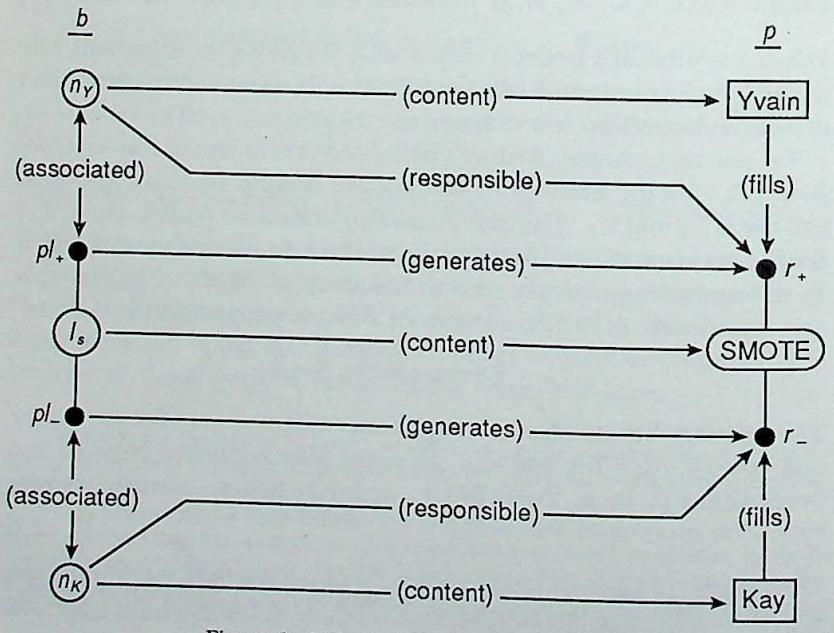


Figure 1: Arthur's belief and its content.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

intensions, and so on. We want our "notions and ideas" to capture what is in common among all these very different models of cognition: there are things shared by different beliefs which explain the internal way in which beliefs must be about the same object or property.

On this theory, one can have two beliefs with exactly the same content or with diametrically opposed contents, such that there is no significant causal relation between them—because they involve different notions. This is a feature of all of the problematic examples that we shall consider. There is nothing particularly puzzling about this—and in fact there is nothing particularly puzzling about any of the examples that we discuss, so long as we simply consider the beliefs and not the reporting of them. Nevertheless, it is a good idea to go over the examples in some detail, for it is these details which our semantic account pays more attention to than others of which we know.

Consider the Prince and the Pauper. Miles Hendon has two notions of Edward Tudor. They have quite different circumstances of origin. One Miles has had for a long time. It is associated (in his beliefs) with his ideas as being a Prince of England, being named 'Edward Tudor', being rich, not being a pauper, not looking like a pauper, not being likely to run into (me) on an average day, and the like. The beliefs with this notion as a constituent influence Miles's behavior when confronted with ordinary sorts of information about Edward Tudor. When he reads an article in the *Times,* for example, it is beliefs with this notion as a constituent which are affected.

His other notion was formed when he saw Edward being set upon by an angry mob—angry because Edward, dressed in rags, had been proclaiming himself to be Prince. This notion is associated with ideas of being out of his mind, being dressed like a pauper, and not being of royal blood. The beliefs involving this notion, and not those involving his old notion of Edward Tudor, influence Miles's behavior toward Edward and Edward's assertions during the period he is associated with him as a comrade, until that point, toward the end of the story, when Miles merges his two notions and comes to believe that Edward the pauper is Edward the Prince.

Perhaps the ultimate doxastic puzzle case is Mark Richard's[13] puzzle about the woman in the telephone booth:

> Consider *A*—a man stipulated to be intelligent, rational, a competent speaker of English, etc.—who both sees a woman, across the street, in a

[13] "Direct Reference and Ascriptions of Belief," *Journal of Philosophical Logic,* XII (1983): 425–452.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

> phone booth, and is speaking to a woman through a phone. He does not realize that the woman to whom he is speaking—*B*, to give her a name —is the woman he sees. He perceives her to be in some danger—a run-away steamroller, say, is bearing down upon her phone booth. *A* waves at the woman; he says nothing into the phone (*ibid.*, p. 439).

The man has two distinct, unlinked notions of the woman. Via one, he believes that she is in danger. This is the notion which arose in virtue of his visual perception of her, and which is associated with an idea of being in grave danger. It is this notion which is involved in the beliefs that motivate his waving out the window. The second notion is an older one, assuming the woman is an old acquaintance. It is associated with an idea of being the person addressed, and not associated with ideas of being the person seen or being in danger. Hence, the beliefs involving this notion do not motivate a warning.

Let us return to Kripke's case. Pierre has the same misfortune as Miles and the man on the phone: he has two notions of the same thing. He has one notion of London which is linked to his memories of the stories and to his use of the word 'Londres'. He has another, unconnected notion of London which is influenced by his perceptions and memories about his present surroundings and which influences his use of the word 'London'. He has a belief associating the former notion with his idea of being pretty, but has no belief associating the latter notion with this idea. In fact, Pierre associates an idea of being ugly with the latter notion.

II.

Our basic idea is simple: a belief report claims that an agent has a belief with a certain content. But the basic idea, unembellished, will not allow us to hold the family of views we want to defend. For (2) and (2′) would claim that Miles Hendon had at least one belief with the content

«Being of royal blood; Edward Tudor»

while (1) and (1′) would deny this—thus contradicting our truth intuitions and the doctrine of opacity.

But our embellishment is also simple. When we report beliefs, there is always some further condition that a belief with the specified content is claimed to meet. The belief report is true, only if a belief meeting that further condition has the right content. What may be novel is our insistence that this additional requirement is part of the proposition expressed by the belief report. Thus, it is a condition on the truth, not merely the felicity, of the report.

Consider (1). In context, (1) provides an explanation of why Miles Hendon did not treat someone he was looking at in a certain way—a way that would have been compulsory for Miles, given the status of

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

that person. We are interested in the content of only those beliefs which motivated Miles's behavior, the beliefs which involve the notion of Edward which arose when Miles saw him being threatened and which explain Miles's treatment of him. The existence of such a notion is clear from the description of the incident. We know that Miles is perceiving Edward and interacting with him on the basis of what he, Miles, perceives. Our view is that, in reporting beliefs, we quite often are talking about such notions, although our belief reports do not explicitly mention them. The general solution to the puzzles is to allow a condition on particular beliefs, over and above a content condition, to be part of the claim made. The version of this strategy we shall pursue here is to take this further condition always to be a specification of the notions that are supposed to be involved in the ascribed belief.

We shall say that a notion that a belief report is about is an *unarticulated constituent* of the content of the report—it is a propositional constituent that is not explicitly mentioned. We shall distinguish another kind of belief report, and say more about the notion of unarticulated constituents in a moment. But first let us see what the semantics of this sort of belief report looks like.

From our account of beliefs, we have the following concepts:

$B(a, b, t)$: b is a belief that belongs to agent a at time t.
Content$(b, t) = p$: p is the content of belief b at time t.
Responsible(n, r, b) : $\exists I, pl$ Associates(b, I, n, pl), and Generates(pl, r).

We take a belief report to be an utterance u of a belief sentence of the form:

A believes that S

where A is a singular term and S is a sentence. We assume a semantics for the use of the embedded sentence, so that $Con(u_S)$ (the content of u_S) is the proposition expressed by the subutterance of u corresponding to S.[14] Where u is a belief report at t which is about notions $n_1, \ldots, n_k$, and $p = Con(u_S)$,

[14] In accord with our simple version of "semantic innocence," we assume throughout that a belief report specifies the content of the ascribed belief by providing a sentence with the same content, as uttered in the report. The puzzle cases that we consider seem to be ones for which this assumption is correct. There are good reasons, however, to think that things do not always work this way. One way of analyzing, "Barbara believes that the Twin Towers are over a foot tall," would involve quantification over contents of beliefs. Other cases of reporting implicit and tacit beliefs might well work similarly. Another case in which a proposition might be "quantified out" is in the use of, "He believes that Russell's yacht is longer than it is." Also, one can use, "Timmy believes that the Tooth Fairy will make him rich," knowing full well that the embedded sentence does not express any proposition (if in fact it does not). These and other cases make us wary of insisting

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

$$Con(u) = \exists b[B(a, b, t) \wedge \text{Content}(b, t) = p \wedge \bigwedge_{r_i \text{ in } p} \text{Responsible}(n_i, r_i, b)]$$

The claim made by the belief report is that the agent a[15] has a belief with the content p, involving the notions $n_1, \ldots, n_k$ (in a certain way).[16] This claim entails the proposition that a has a belief with the content p, but the truth of that proposition is not sufficient for the truth of the report—the report says more than that about the ascribed belief.

We shall say in such cases that the notions that the belief report is about are *provided* by the utterance and its context. Note that the provided constituents of the report's content are not existentially quantified.

Let us see how this theory works with Miles, Edward, and our intelligent reader. We take our reader to be talking about n_{vis}, the notion Miles acquires of Edward from visually perceiving him on the occasion of the rescue. $Con(u_S)$ is just the proposition ≪Being of royal blood; Edward≫. So our reader is saying with (1) that there is no belief that associates Miles's idea of the property of being of royal blood with Miles's notion n_{vis}. He is not contradicting any proposition that Miles has some other notion of Edward Tudor which is so associated.

---

that a content proposition is always specified in a belief report. The present strategy can be extended in relatively simple ways to account for such cases.

[15] Yet another simplification: we ignore the fact the many uses of singular terms, including terms in the subject position of belief sentences, are not directly referential. "Attributive" uses of definite descriptions really should be handled differently. Note also that we really should treat the idea in a belief in the same way we treat notions here; though the puzzles considered here do not turn on this, others certainly do.

[16] Here one major difference from the "official" belief-report semantics in ch. 10 of Perry's *Situations and Attitudes* (p. 256) is apparent. There, a belief report is true if the agent has any belief with the specified content. There is a further crucial difference that is not so obvious. Barwise and Perry countenance beliefs as real, concrete things, as we do here. But these beliefs are represented as situations of an agent being related to an anchored belief schema. Belief schemas are abstract objects in which what we have called notions are represented by indeterminates. Although the way this all works is quite complicated, in the end beliefs are individuated by belief schemas—abstract objects—and the things in the world to which the indeterminates in the schemas are anchored. But indeterminates are not notions, and, we think, relations to anchored belief schemas are not quite fine-grained enough to individuate beliefs in the ways needed for belief reports. So we suggest two major changes to the account in *Situations and Attitudes:* we give ourselves the theoretical machinery to talk about notions and ideas directly; we then claim that these things are among the subject matter of belief reports (via the mechanism of unarticulated constituents), and are not merely quantified over.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

And, in fact, a proposition of this latter kind might be just what our reader intends to claim with (2). Imagine the case in which he reads that Miles Hendon is shouting, while treating Edward as a mad fool, "Prince Edward is a man of royal blood, you fool, who would not dress in rags." Our reader might intend to say, of the notion involved in the beliefs that motivate this behavior, that it both is of Edward Tudor and is associated with the idea of being of royal blood.

If so, our reader would surely be consistent, direct, and innocent. On the one hand, the proposition he in turn denies and affirms Miles's belief in is just the singular proposition that contemporary theories of direct reference assign to the utterances of "Edward Tudor is of royal blood" and "He is of royal blood" in the described contexts. On the other, the denial and affirmation are completely consistent.

III.

We have claimed that, in belief reports, an n-ary relation is reported with an n-minus-one-place predicate. On our account, the complex relation invoked in belief reports is a four-place relation: an agent believes a proposition at a time relative to a sequence of notions. But there is no argument place in the 'believes' predicate for the sequence of notions. The notions are unarticulated constituents of the content of the report.

Propositions have constituents. The proposition that Yvain smote Kay has Kay as a constituent—Kay himself is in that claim. When Arthur says, "Yvain smote Kay," there is no great mystery about why Kay, rather than someone else, is part of the claim Arthur makes: Arthur uses the name 'Kay', which, as he uses it, refers to Kay. Kay is the content of Arthur's utterance of 'Kay'. This is what it is to be an articulated constituent of the content of a statement.

It is very common in natural languages for a statement to exploit unarticulated constituents. When we consider the conditions under which such a statement is true, we find it expresses a proposition that has more constituents in it than can be traced to expressions in the sentence that was spoken. Each constituent of the content which is not itself the content of some expression in the sentence is an unarticulated constituent of the content of the statement.

We report the weather, for example, as if raining and snowing and sleeting and dark of night were properties of times, but they are one and all relations between times and places. If I say, "it is raining," you understand me as claiming that it rains at that time at some place the context supplies. It often is, but need not be, the place of utterance. If I am talking to a friend in Kansas City on the phone, or watching news reports about the continuing floods in Berkeley, you

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

may understand me to be talking about those places rather than the place where we both are.

The phenomenon of unarticulated constituency is similar to that of indexicality in the reliance on context. But the two phenomena should not be conflated. If we say, "It's raining *here*," an expression in our statement identifies the place. The place is articulated in a context-sensitive way. In the case of indexicals, expression and context share in the job of identifying the constituent, according to the conventional meaning or character of the indexical. In a case of underarticulation, there is no expression to determine the constituent in this way.

It would be misleading, however, to say that, in the case of unarticulated constituents, the context alone does the job. The whole utterance—the context and the words uttered—is relevant to identifying the unarticulated constituent. Thus, a change in wording can affect the unarticulated constituent, even though it is not a change in an expression that designates that constituent. Suppose I am in Palo Alto talking on the phone to someone in London; we both know that it is morning in Palo Alto and evening in London. If I say, "It's exactly 11 A.M.," I will be taken to be talking about the time in Palo Alto; if I had said, in the same context, "It's exactly 8 P.M.," I would be taken to be talking about the time in London.

The important principle to be learned is that a change in wording can precipitate a change in propositional constituents, even when the words do not stand for the constituents.

Unarticulated constituency is one example of the incrementality of language. In the circumstances of an utterance, there always is a great deal of common knowledge and mutual expectation that can and must be exploited if communication is to take place. It is the function of the expression uttered to provide just the last bit of information needed by the hearer to ascertain the intended claim, exploiting this rich background. What is obvious in context we do not belabor in syntax—we do not articulate it.

This is by no means to transgess the intuition of the systematicity of language which is commonly reflected in principles of "compositionality." Since we finite creatures are able to make and understand a potential infinity of claims, there must be systematic features of our statements which explain our infinite abilities in something like a combinatorial fashion—in terms of our more finite abilities to understand the contributions of specific features of statements toward the claims made. But there is no reason to assume that these features of statements must all involve syntactic expressions. It is just as systematic for a form of speech, like a belief report or a report of

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

rain, to *call for* a propositional constituent that meets, say, certain conditions of relevance and salience, as it is for a form of speech to have a syntactic expression *stand for* a propositional constituent.[17]

Consider our practices of reporting velocity. A claim that an object is moving at a certain velocity makes sense only if it is understood with respect to what the velocity is to be assessed. We say that velocity is relative to an observer, or a frame of reference—we must count something as stationary. But we articulate this additional parameter of velocity claims only when it is not obvious what is to count as stationary. We have in English a number of general-purpose constructions for articulating commonly suppressed constituents of a claim. We say, 'with respect to . . .' or 'relative to . . .' or 'in the sense that . . .'. The more likely the unarticulated constituent is to be unclear, the more likely it is that we have a natural way to articulate it.

In the case of belief reports, in which notions are unarticulated, we do have rough and ready ways to clarify just which notions we mean to talk about. We say, for instance, that Miles believes that Edward is a peasant in one way—the way related to the boy in front of him, not in the way related to the Prince. Or we add to the report, "that is, he thinks the boy in front of him, who really is Edward, is a peasant." Or we specify how Miles would or would not "put" his belief. Or we allude to the evidence which led Miles to form the belief, or to the actions it would be likely to bring about. Each of these devices can succeed in distinguishing among the two notions which in context can seem equally relevant, thus eliminating possible confusion about which notion we mean to talk about.

We do not, of course, have a very direct way of specifying the notions we mean to talk about in belief reports. This is due to the fact that it almost always obvious which notion a speaker is talking about. Where it is not, we either use one of the devices just mentioned, or leave the language of belief reporting altogether and talk instead about what the agent would say or would do.

IV.

Unarticulated constituency and direct reference are of a single stripe. In fact, if we take the term 'reference' in the ordinary sense in which it does not require a referring expression, unarticulated constituency can be seen to result from a kind of direct reference—perhaps "tacit" reference. When a speaker claims that "it's raining," she

[17] For more on unarticulated constituents, see Perry, "Thought without Representation," *Proceedings of the Aristotelian Society*, suppl. vol. LX (1986): 263–283. There, a systematic semantics for some underarticulated constructions is given, which is connected to a recursive model of syntax in the usual way.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

is referring to a place and not to a description of, nor a condition on, a place. In the same way, on our view, a belief reporter refers to an agent's notions. We have chosen not to talk this way in our official account only to avoid being read as claiming that notions are referred to by the reporter's words.

A difficult issue facing all views of direct reference, and ours in particular, is the need to make sense of intuitions about truth and falsity in cases of reference failure. This problem is especially acute for our account in some cases of denials. Consider the following example. A blind man is facing in the direction of a distant building. Someone, unaware of the man's blindness, says, "He believes that that building is far away." One normally would take this report to be about the notion the man has as a result of his current visual perception of the building. The speaker is trying to refer (though not with a word) to such a notion, to provide such a notion for the report to be about. But of course there is no such notion in this case. Is this report false, or, owing to a failure of tacit reference, does it fail to express a proposition? Certainly, we ordinarily would respond not by saying, "You have failed to express a proposition," but "He doesn't believe that"—and we have the strong intuition that this denial would be true.

Compare the following case: an astronaut on the moon[18] says, "It's three o'clock." Typically, this sentence would be used to express the claim that it is three o'clock in *Z*, where *Z* is the time zone in which the utterance takes place. The confused astronaut thinks that there are time zones on the moon, and he intends to claim that it is three o'clock in "Z," which is the time zone he is in. But there is no such time zone. So he fails to express a proposition. We feel no qualms, however, about denying his claim: "It's not three o'clock. There are no time zones on the moon, you . . . ."

The present difficulties are often discussed in connection with "negative existential" claims. But the same issues arise with respect to all sorts of denials in which the speaker believes there to be reference failure. A child who sincerely asserts, "Santa will come tonight," fails to refer, and therefore, on most direct reference accounts, fails to express a proposition. But the parent who responds, "Santa will not come tonight," explaining that there is no Santa, makes what seems to be a true claim, despite the fact that the use of 'Santa' does not refer.

Note that these examples would present no trouble for descriptional theories of reference. For if in these cases the original

[18] John Etchemendy brought up this version of Wittgenstein's example.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

speakers are seen, not as attempting to provide a specific thing to be a propositional constituent, but merely as claiming that there is a thing meeting a certain condition (being the generous elf known as 'Santa', being the local time zone, or being the man's perceptual notion of the building), then the claims are straightforwardly false and the denials are true.

The descriptional theories have even more than this kind of extensional correctness going for them; it is because the cited conditions—call them *providing conditions*—are not satisfied that the denials are true. In the child's use of 'Santa', the providing condition, of being the generous elf known as 'Santa', plays a central semantic role, even though it is not the referent of the child's use of the name. It is a condition that the child expects to be filled as a precondition of successful reference. He expects to refer successfully to a thing in virtue of its meeting the providing condition. His supposed ability to refer to a thing by using the name 'Santa' depends on the condition's being satisfied. Similarly, the astronaut takes it that he can talk directly about a time zone, that he can provide one, because it meets the providing condition of being the local time zone. And, we claim, the belief reporter expects to be able to talk directly about a notion because it satisfies the condition of being the man's perceptual notion of the building.[19]

A normal, successful case of direct reference involves a speaker referring to an object in virtue of that object's satisfying a providing condition. Reference failure involves failure of a presupposition, namely, the presupposition that a providing condition is satisfied. Now, expressions like proper names and underarticulated phrases that normally invoke devices of direct reference are sometimes used where there is no presupposition that the relevant providing conditions are satisfied. The denials in the cases of the blind man, the astronaut, and Santa are like this. In each of these denials, the speaker does not presuppose that there is a thing meeting the providing condition that is invoked by the utterance. Instead, we claim, the speaker *raises the providing condition to constituency*—he talks about the condition itself rather than about a supposed thing that meets it. The providing condition now plays a semantic role—as a

[19] Just which providing conditions are invoked in a given case depends on a wide range of circumstances. Also, there usually is more than one such condition for a given use of a term. Providing conditions for a use u of 'here' by speaker A at location l, for instance, include the conditions of being the location of the utterer of u, being where A is, and being l. In the 'Santa' case, we have the conditions of being the referent of the utterance of 'Santa', being the relevant thing known as 'Santa', being the generous elf known as 'Santa', and so on.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

constituent of the proposition expressed in the denial—more central than its usual auxilliary role of providing a propositional constituent.

In particular, the claim expressed by 'Santa will not come tonight' (in the described circumstances)[20] is to the effect that there is no generous elf known as "Santa" who will come tonight. And the proposition expressed by 'It is not three o'clock' is that there is no local time zone such that it is three o'clock there. And the content of 'He does not believe that that building is far away' is the claim that there is no perceptual notion of the building such that the man has a belief involving that notion, with the content that the building is far away. The denials are thus true, and their truth is consistent with our claim that the assertions they deny strictly speaking fail to make claims.

Of course, for each of the original, claimless assertions, there is a proposition closely related to the kind of proposition the speaker intends to express, which we can for most purposes charitably treat as the content of the statement. Specifically, we can take the speaker to have expressed the claim that there is a thing meeting the invoked providing condition, such that so-and-so. In fact, the speaker of such an assertion is pre-assertively committed to this proposition, in virtue of his commitment to the presuppositions that must be satisfied if he is to make a successful claim in the way he intends.

Above we analyzed our reader's utterance of (2) in an imaginary case in which Miles has been shouting about the Prince. In fact, Miles was not shouting, "Edward Tudor is of royal blood," at the time he encountered the boy. The reader actually has no specific actions on Miles's part to which he can tie such a notion of Prince Edward. It is obvious from the general tenor of the novel, however, that Miles would have such a notion. Every full-witted adult in England at the time has a notion of Prince Edward—one they acquired shortly after he was born which motivates their behavior in regard to the Prince of Wales, such as their use of the phrase 'Prince Edward', their decorum when the royal procession goes by, and the like. Our reader may not be able to pick out anything very specific in Miles's behavior to serve as evidence that he has such a normal notion of the Prince. But he has every right to suppose that he has one.[21]

[20] In circumstances where it is presupposed that the providing condition is met, the denial expresses just the negation of the proposition (if there is one) expressed by the corresponding assertion.

[21] What counts as being a normal notion certainly depends, we think, not only on what is common in a community, but also on other aspects of the background of the discourse, including facts about what is relevant to the goals of the discourse. We would expect an account of "being a normal notion" to exhibit many of the same

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

It may seem implausible to suppose that our reader, in using (2), can directly provide a notion for the report to be about, since the reader is not directly acquainted with such a notion. If this intuition is right—an assumption we shall question in a minute—our machinery gives us a natural way of respecting it: this is a case in which, instead of a notion, a providing condition becomes a propositional constituent. What our reader is claiming with (2) is that there is some normal notion via which Miles believes that Edward is royal; that is, the condition of being a normal notion of the Prince is the unarticulated constituent. The report, on this construal, is an example of a second kind of belief report—in which notions are not provided, but instead are constrained by provided conditions; the report is about those conditions in the sense of "about" appropriate to propositional constituents.

For this (supposed) second kind of belief report we can give the following account. Where u is a belief report at t which is about conditions $C_1, \ldots, C_k$, and $p = Con(u_S)$, where u_S is the subutterance of u corresponding to the object sentence S,

$$Con(u) = \exists b[B(a, b, t) \wedge \text{Content}(b, t) = p \wedge \exists n_1, \ldots, n_k \bigwedge_{r_i \text{ in } p} (C_i(n_i) \wedge \text{Responsible}(n_i, r_i, b))]$$

So we have room in our framework for two sorts of belief report, corresponding to whether notions are themselves *provided* or merely *constrained* by conditions. Supposing for now that there really are two kinds of belief report, how can we know, for a given report, of which kind it is? One way, surely, is to look at what would happen if the appropriate notions were to fail to exist. If the report would then be false, then it is a case of notion constraint rather than provision; if the report would fail to make a claim, then it is a case of (attempted) notion provision.

Of course, we have seen how, in a case where an attempt to provide a notion fails, a proposition closely related to what the speaker is trying to express takes center stage. This is the false proposition to the effect that the agent has a notion which meets the invoked providing condition and which is involved in a belief with such-and-such content. Given this fact, our intuitions about whether a belief report fails to make a claim or is simply false are in the same boat as our intuitions about the truth value of the child's claim that Santa is

---

features as an appropriate account of "knowing who b is," which certainly is background-sensitive in many ways. See, for example, Böer and Lycan, *Knowing Who* (Cambridge: MIT, 1985).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

coming. The falsity of the closely related propositions, plus the truth of the natural denials of these statements, may well obscure intuitions about the truth of the original claims.

In this paper, we adopt officially the position that there really are belief reports of the second kind (which are about conditions rather than notions). Given our points about providing conditions and propositions to which speakers are pre-assertively committed in cases of direct reference, however, a plausible case can be mounted for the view that, in all successful belief reports, specific notions are provided for the report to be about.[22]

Assuming, now, that there are two classes of belief reports, there is no reason to suspect that all reports will fall clearly into one camp or the other. For example, if our reader simply assumes that Miles must have a normal notion of King Henry and expects his audience to do the same, then it makes little difference whether he claims that Miles has a belief involving that notion (notion provision) or just a belief involving a normal notion (notion constraint). Since it makes little difference, our reader need not go to any pains to indicate which of the claims he is making; his report simply can land between the two claims.[23]

V.

With that out of the way, let us turn to our examples.

First, a recap of the semantics of (1) and (2). We shall treat (1) as a case of notion provision. The provided notion is Miles's notion of Edward which is connected with his perception of and actions toward Edward in the mob incident. The reader claims that Miles does not have a belief involving that notion, with the content that Edward is of royal blood.

With (2), the reader provides a condition on notions, the condition of being a normal notion of Prince Edward. The reader claims that Miles has a belief involving some normal notion of Edward, with the content that Edward is of royal blood.

In the Pierre case, the sentence (3) gets used in two reports, first in a discussion of Pierre's initial acquaintance with London through stories, then in a discussion about Pierre's thoughts of his adopted

[22] This view is argued in Crimmins, *Talk about Beliefs*, though in the end the argument rests precariously on the fact that none of the examples considered as natural candidates for reports of the "second kind" seems clearly to be as required.

[23] There is another way, also, in which a report can land between the notion-provision and the notion-constraint types of report. It is not hard to concoct cases in which one notion is provided and another is constrained; a natural construal of our reader's report "Miles believed that he (Edward in rags) was less noble than Prince Edward" might go along these lines. So in the general case, both notions and conditions may be provided; there are no difficulties in formalizing this along the lines already given for the pure notion-provision and pure notion-constraint analyses.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

home. Call these reports u_3 and u'_3. Pierre actually has two notions of London, one relevant to each discussion; call the first n and the second n'. The notion n meets the condition C of being a notion germane to the discussion of Pierre's reaction to the stories; the notion n' meets the condition C' of being a notion germane to the discussion of Pierre's new home.

If one of the two analyses is uniquely correct for u_3 and u'_3, it is perhaps the account in terms of notion constraint. The speaker of the former report is claiming that Pierre has a belief involving some notion germane to the current conversation about the stories, with the content that London is pretty. The speaker of the latter report requires that the belief involve some notion relevant to the conversation about Pierre's new home.

If the circumstances of u_3 and u'_3 are such as to make the notions n and n' clear and present to the speakers and their audiences, then the analysis should be in terms of notion provision. If this is the case, then the speaker of u_3 claims that Pierre has a belief involving the notion n with the content that London is pretty; the speaker of u'_3 claims that Pierre has a belief involving the notion n' with the content that London is pretty. If the circumstances of the two reports are less clear-cut, as they often are, then, as noted earlier, the claims made by the speakers might fall between those offered by the notion-provision and notion-constraint accounts. There just might be no saying.

Note, though, that any of these analyses constitutes a solution to the puzzle. The claim made in u_3 is simply true, and the claim made in u'_3 is simply false.

Kripke presents the puzzle as arising from a few very plausible principles about belief reports, including:

> Disquotation: If a normal English speaker, on reflection, sincerely assents to 'p', then he believes that p (*op. cit.*, pp. 248/9).
>
> Translation: If a sentence of one language expresses a truth in that language, then any translation of it into any other language also expresses a truth (*op. cit.*, p. 250).

On our account of belief reporting, neither of these principles is at all plausible in general. Each principle presupposes that it is belief sentences that are true or false. On our view, a single sentence, like (3), can be used in both true and false reports. Kripke assumes that, because of the lack of obviously context-sensitive words, (3) can be considered more or less "eternal." But words are not the only sources of context-sensitivity; the presence of unarticulated constituents also can widen the gap between a sentence and the proposition expressed by a statement of it. And that is what happens in the Pierre case.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Richard lists three sentences considered as uttered by A watching B in the phone booth:

(4) I believe she is in danger.
(5) I believe you are in danger.
(6) The man watching you believes you are in danger.

A uses (4), clearly, to make a true report. His notion n_{vis} of B which stems from his view out the window, which is associated with his idea of being in peril, and which causes his waving is supplied. It is claimed that A has a belief involving n_{vis} with the content that B is in danger. He in fact has such a belief.

The man would not sincerely use (5) over the phone; if sincere, he certainly would deny (5). The natural intuition, we think, is that a use of (5) in the described circumstances would make a false claim. (It is this reaction which Richard sets out to prove mistaken. The very possibility of our semantics shows that his proof is in error.)

The set-up for (6) is as follows. B sees a man, A, in a building across the street waving frantically. Amused, she says (over the phone), "the man watching me believes that I'm in danger." Echoing her, A utters (6). Surely, B's claim is true. And if so, A's use of (6), which is in explicit agreement with her, is true also.

So we hold that the use of (5) is false while that of (6) is true. But how can this be? The two reports are uttered by the same person in the same circumstances, they ascribe beliefs to the same agent, and they use precisely the same embedded sentence, understood in the same way! The only difference is the way in which the man is referred to—in the one case with 'I', in the other with 'the man watching you'.

Difference enough, we think. The pragmatic principle of self-ascription applies to (5) but not to (6):

> Self-Ascription: An utterance of 'I believe that . . . τ . . .' provides (or, is about) the notion that is connected to the speaker's use of 'τ'.

Using 'I' in (5), A thus directs attention to the notion (n_{phone}) that is linked to his use, in (5), of 'you'—the notion of B which is associated with his idea of being the one he is addressing[24] and not associated with his idea of being in danger. So A's use of (5) makes the claim that he has a belief involving n_{phone}, which has the content that B is in danger. He in fact has no such belief.

In (6), A is discussing those beliefs of the man watching B, that is, of A himself, which explain the frantic gestures directed at B. So he claims that the man has a belief involving n_{vis}, the notion linked to his perception out the window and his gestures of warning, which has the content that B is in danger. In fact, A has such a belief.

[24] More precisely, with his idea I_{addr}, the idea that has the context sensitive semantic role picking out the person being addressed.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Richard's case is especially interesting because it shows how a contextual shift can be brought about by a change in wording outside of the embedded sentence in a belief report. This gives added force to our analysis of substitution worries: the wording changes in the usual cases of reluctance to substitute are responsible, not for changes in meaning or explicitly specified content, but for changes in what is provided by context for the reports to be about.

Our semantics allows that, for a given belief sentence, absolutely any of the agent's notions may be provided—there is no semantic restriction on what notions may be provided in a use of a given sentence. But there are many pragmatic principles, like self-ascription, that constrain which notions can be provided in the normal case. It is semantically but not pragmatically possible for a use of 'I believe I am not me' or (normally) '*S*, but I do not believe that *S*' to be true. Although it is semantically possible, in W. V. O. Quine's example, for an utterance of 'Tom believes that Cicero is Tully' to express a true proposition (say, if Tom's 'Cicero' notion is provided twice over), there may be no very natural use of that sentence which in fact expresses the proposition (although surely we can concoct a Richard-ish example to put this point in doubt). In the normal case, the use of different names for Cicero serves as a strong, though perhaps defeasible, indication that the names have some importance to what's being said over and above just standing for Cicero. Such a difference in names requires a sufficient reason—in this case, a difference in which notions are being provided to play the corresponding roles in the ascribed belief.

VI.

The relation of the present proposal to Fregean semantics for belief reports should be relatively clear. The broad similarity consists in the agreement that a belief report specifies, in addition to simply which objects the agent is claimed to have a belief about, also just how the agent is cognitively connected to those objects. On our account, the report specifies (or constrains) the particular notions allegedly involved in the belief. On a Fregean account, "senses" are specified.

Two crucial differences separate the accounts. First, we stress the particularity and unsharability of notions. Since notions are full-fledged particulars immersed in the causal order, they have a great array of different features that we can exploit to provide them in our belief reports. They are involved in beliefs, associated (sometimes) with words, formed in specific circumstances, connected to perceptual situations, reasonings, and actions; they survive the formation and abandonment of beliefs in which they are involved; and so on. We can use each of these kinds of fact to give us a handle on a notion, a way of picking it out. This frees us from a problem often noted about the Fregean strategy: it appears that, on most natural con-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

struals of what senses are, we often do not know just what sense an agent attaches to an object (we do not grasp it), and so we cannot know just what we are attributing to the agent with a belief report, which, after all, must be about the agent's ways of thinking.

As we have said, there is nothing in our view incompatible with something like Fregean senses, considered as entities which we can use to classify an agent's notions. A Fregean might well take our talk of "notions" as an account of what it takes for an agent to grasp a sense—agents grasp senses in virtue of having appropriate notions.

The second departure from a Fregean account is in our claim that the agent's ways of thinking about things (her notions), though they are specified in a belief report, are not the referents of the words occurring in the embedded sentence. This difference becomes especially important in the analysis of certain kinds of reports: those with content sentences containing devices of underarticulation, and those with no content sentences at all, but which instead are completed with the likes of 'what you said', 'the same thing', and 'Church's Thesis'.

VII.

The account of belief reports sketched here closes some doors. If, as we claim, a single belief sentence can be used in both true and false reports, then there can be no simple logic of such sentences. The simplest possible rule,

$$\frac{A \text{ believes that } S}{A \text{ believes that } S}$$

does not hold in general, as we learn from Kripke's puzzle.

Even a logic of belief sentences restricted to a single context will prove difficult.[25] Although a relativized version of the above rule will certainly hold, this one,

$$\frac{\begin{array}{ll} A \text{ believes that } S & (\text{relative to } c) \\ A = B & (\text{relative to } c) \end{array}}{B \text{ believes that } S \quad (\text{relative to } c)}$$

will not, as we learn from Richard's puzzle.

Also closed is the prospect of a strictly compositional semantics for belief sentences. The semantic values of the subexpressions in a belief report, on our analysis, do not provide all the materials for the semantic value of the report itself. Notions and conditions on no-

[25] Here we mean 'context' in a sense such that various different statements can be made in the same context. One way of taking our claims in this paper would be as denying the general usefulness in semantics of such a restricted notion of context. Taken this way, we have claimed that such things as the words used in a statement can affect the semantically relevant parts of the statement's context.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tions are not articulated, but end up in the contents of reports; so the semantics of belief reports is in an important way noncompositional.

In addition, our account denies what some have seen as a primary desideratum for theories of belief: that a belief report claims simply that a binary relation holds between an agent and an object of belief.

And, perhaps worst of all, we have given an account on which it appears to be next to impossible to give a complete, systematic account of which claims are made by which belief reports. We have claimed that belief reports are context-sensitive, that they invoke unarticulated constituents, without offering any general method for determining what the relevant contextual factors are, and how they give rise to these unarticulated constituents of belief reports.

Tempted as we are to view each of the above results as an insight rather than a drawback, we realize that we have abandoned many of the issues and goals commonly pursued in this area. But we think the account opens many doors as well.

Whereas there is little possibility of an interesting logic of belief sentences, the logic of beliefs, notions, and ideas is available. Such issues as logical and analytic closure of belief, explicit versus implicit belief, and inferential issues in belief change really belong to the logic of beliefs rather than to the logic of belief sentences. We can explore the logic of the relations we have seen as underlying our ordinary talk about beliefs—but this logic will not be a logic of ordinary language.

Of course, we have explained very little about what beliefs, notions, and ideas are. But we think our partial account of them raises obvious questions in theories of representation, action, perception, and the metaphysics of mind.

Our semantics is not compositional, but there is system in the noncompositional mayhem. The ways in which notions and conditions on notions are provided have yet to be explored to any great extent. But the discussions of the belief puzzles suggest several directions from which to look at these mechanisms.

Last, the move to unarticulated constituents emphasizes the importance of pragmatic facts about language to the study of what seem like purely semantic issues. In order to express claims, we exploit a tremendous variety of facts, conventions, and circumstances, of which the meanings and referents of our terms form just a part. So it is a mistake to relegate pragmatics to matters of felicity and implicature. In the case of belief reports, it is central to understanding content and truth.

MARK CRIMMINS

Cornell University

JOHN PERRY

Stanford University

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

SOCIAL POWER AND HUMAN AGENCY*

MY project is this: from among the varied relationships between people, how can we distinguish those in which one person exercises power over another? I shall not claim to have a positive answer to this question, but I shall claim that three suggested answers do not work.

From the class of cases in which one person, *A*, affects another, *B*, these three suggestions pick out the cases involving the exercise of power in terms of:

(1) *A*'s affect on *B*'s interests.
(2) *A*'s wants or intentions in affecting *B*.
(3) *A*'s responsibility for affecting *B*.

The three definitions have in common an appeal to various aspects of human agency, to *A*'s wants or responsibility, or to *B*'s interests. Very roughly, the problem they all share is an inability to comprehend adequately forms of social power which interfere with human agency, forms of power exercised over people's wants, intentions, and perceived interests.

There is little agreement on the details of classifying forms of social power, but one broad scheme is that of J. K. Galbraith.[1] He classifies power into *condign* power, which includes physical force and negative sanction, *compensatory* power, which includes inducement and incentive, and *conditioned* power, which he describes as follows:

> While condign and compensatory power are visible and objective, conditioned power, in contrast, is subjective; neither those exercising it nor those subject to it need always be aware that it is being exerted. The acceptance of authority, the submission to the will of others, becomes the higher preference of those submitting. This preference can be deliberately cultivated—by persuasion or education. This is explicit conditioning. Or it can be dictated by the culture itself; the submission is considered to be normal, proper or traditionally correct. This is implicit conditioning (*ibid.*, p. 24).

What Galbraith calls conditioned power affects not only the final outcome of a person's choices, but also the process of choosing itself,

* I would like to acknowledge the financial support of the Social Sciences and Humanities Research Council of Canada and the hospitality of the University of Toronto Philosophy Department during the period in which this essay was written.

[1] *The Anatomy of Power* (Boston: Houghton Miflin, 1983).

0022-362X/89/8612/712–726 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the site of human agency.[2] Galbraith's concept of "implicit conditioning" points to why a critique of the human-agency view of power is interesting. If the agency view is correct, only agents (human beings or perhaps corporate bodies) can have power. Wants, intentions, responsibility, or interests can be attributed only metaphorically to the "culture itself," or other social structures. The agency view systematically misses the power of social structures to mold human personality. It could not comprehend, for example, mechanisms of gender formation.

My first point is simple: you cannot define the concept of power in terms of its effects on people's real interests, and then characterize real interests in terms of their autonomy from the effects of power. The point of a definition of the concept of power is to make possible the identification of power in its many guises, and a circular definition is not at all helpful. My moral is also very simple: power should be defined independently of the interests of the dominated, real or otherwise.

Who would define power in terms of real interests, and why? The philosophical sociologist Steven Lukes[3] does, and he does so for admirable reasons. Lukes is interested in analyzing the power of one social actor, *A*, over another, *B*, rather than a social actor's power per se. He is interested in *A*'s power over *B*, because he is interested in ascertaining the responsibility of one actor for his, her, or its effect on another. (A social actor, for Lukes, can be a collectivity.) "The point," he writes, "of locating power is to fix responsibility for consequences held to flow from the action, or inaction, of certain specifiable agents" (*ibid.*, p. 56). The problem is to specify which cases of *A* affecting *B* are relevant to the power of *A* over *B*. I may startle you when I sneeze, but I have not thereby exercised power over you. Lukes specifies which affectings are significant to the existence of power in terms of *B*'s interests: "*A* exercises power over *B* when *A* affects *B* in a manner contrary to *B*'s interests" (*ibid.*, p. 27).

[2] Rousseau said, "[t]he most absolute authority is that which penetrates into a man's inmost being, and concerns itself no less with his will than with his actions." *A Discourse on Political Economy*, Cole, trans., cited in J. P. Day, "Threats. Offers, Law, Opinion and Liberty," *American Philosophical Quarterly*, XIV (1977), p. 263.

[3] *Power: A Radical View* (London: Macmillan, 1974). Lukes's work has been much discussed; some interesting criticisms are: Alan Bradshaw, "A Critique of Steven Lukes' 'Power: A Radical View'," *Sociology*, X (1976): 121–127, which is followed by Lukes's reply; Ted Benton, " 'Objective' Interests and the Sociology of Power," *Sociology*, XV (1981): 161–184; and David West, "Power and Formation: New Foundations for a Radical Concept of Power," *Inquiry*, XXX (1987): 137–154.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

To complete his account, he must explicate 'contrary to *B*'s interests'.

Lukes believes that the various views of power come apart in their conceptions of interests (*ibid.*, pp. 34–35). The liberal, or one-dimensional, view of the American pluralists[4] looks to interests as revealed in actual conflicts. The reformist, or two-dimensional, view of the critics of pluralism[5] looks to discernable interests which are, however, kept out of the arena of conflict. But the radical, or three-dimensional, view of power, the one that Lukes is advocating, points out that the perceived interests of subservient social actors may themselves be produced by the power of dominant social actors. He writes:

> To put the matter sharply, A may exercise power over B by getting him to do what he does not want to do, but he also exercises power over him by influencing, shaping or determining his very wants. Indeed, is it not the supreme exercise of power to get another or others to have the desires you want them to have—that is, to secure their compliance by controlling their thoughts and desires (*op. cit.*, p. 23)?

It is important to Lukes, as it should be to any social theorist, to give an account of power which comprehends power exercised through the manipulation of perceived interests.

To be able to identify the manipulation of subjective interests, we need some baseline against which we can check apparent interests for the distortions of power. In other words, we must be able to identify people's real interests. But how can we do this in a way that avoids paternalism and the diabolic reasoning that leads to totalitarian tyranny? Lukes's desire to avoid paternalism leads him to explicate *B*'s real interests as the interests *B* herself would choose under certain conditions, that is, when *B* can choose autonomously. There is "an empirical basis for identifying real interests," Lukes claims, when

> The identification of these is not up to A, *but to B,* exercising choice under conditions of relative autonomy and, in particular, independently of A's power—e.g. through democratic participation (*op. cit.*, p. 33).

Although Lukes's discussion here is sketchy, I think it is fair to make the following conclusion from this passage. *B*'s real interests are

[4] As exemplified in R. A. Dahl's *Who Governs: Democracy and Power in an American City* (New Haven: Yale, 1961).

[5] Importantly, Peter Bachrach and Morton Baratz, "The Two Faces of Power," *American Political Science Review,* LVI (1962): 947–952, and "Decisions and Nondecisions: An Analytic Framework," *American Political Science Review,* LVII (1963): 641–651.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

those *B* would choose if, and only if, *B* were to choose independently of both *A*'s power and everyone else's power. The relevant aspect of power is power over *B*'s choice of interests. In other words, *B*'s real interests are those interests *B* would choose to have if, and only if, *A* had no power over *B*'s choice of interests and no one else had power over *B*'s choice of interests.[6]

The problem for Lukes's account of power arises when we try to put his specification of real interests together with his definition of power. The latter is:

> I have defined the concept of power by saying that A exercises power over B when A affects B in a manner contrary to B's interests (*op. cit.*, p. 34).

We have to be clear about what aspect of *B* we are talking about. If it is *A*'s power over *B*'s actions in which we are interested, there is no immediate problem. Putting the two definitions together we can say:

(1) *A* has power over *B's actions* when *A* affects *B's actions* in a manner contrary to the interests *B* would choose if, and only if, *A* had no power over *B*'s choice of interests and no one else had power over *B*'s choice of interests.

A's power over *B*'s actions is defined in terms of *A*'s power over *B*'s choice of interests. But what happens when we try to identify when *A* has power over *B*'s choice of interests? This identification has to be made in order to define 'power' fully. Definition (1) is not complete as it stands. It must be supplemented with a definition of *A*'s power over *B*'s choice of interests.

(2) *A* has power over *B's choice of interests* when *A* affects *B's choice of interests* in a manner contrary to the interests *B* would choose if, and only if, *A* had no power over *B*'s choice of interests and no one else had power over *B*'s choice of interests.

This definition tells us that, in order to identify when *A* has power over *B*'s choice of interests, we must be able to identify when *A* does not have power over *B*'s choice of interests. But we cannot identify the absence of power without first being able to identify power. Definition (2) is not, however, blatantly circular. *A*'s power over *B*'s choice of interests in our actual situation is defined in terms of *A*'s lack of power over *B*'s choice of interests in another, possible yet

[6] Bradshaw notes that Lukes's situation of relative autonomy is a "ridiculously barren, social arena" (*op. cit.*, p. 122). For example, people cannot avoid being infants at some time in their lives, and must be always socialized through the power of others. A better definition of 'relative autonomy' would take 'relative' to heart and talk of minimizing power rather than remove it.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

contrary-to-fact, situation. But the definition is useless because it gives us no way of identifying A's power or lack thereof in the counterfactual situation. All we can do is reapply definition (2) in the counterfactual situation and thereby generate the problem again. Presumably 'A does not have power over B' means that A does not affect B in a manner contrary to B's interests. Substituting for *A had no power over B's choice of interests* in (2), in a vain attempt to eliminate the concept of power from its own definition, we would get:

(2′) A has power over B's choice of interests when A affects B's choice of interests in a manner contrary to the interests B would choose if, and only if, A did not affect B's choice of interests in a manner contrary to the interests B would choose if, and only if, *A did not have power over B's choice of interests and no one else had power over B's choice of interests* and no one else had power over B's choice of interests.

The undefined concept of power just keeps reoccurring as we try to eliminate it. This leaves us pursuing the identification of A's power over B's choice of interests through an endless series of counterfactual situations.

I think Lukes gets to this impasse for the very best of reasons. He is exactly right when he claims that an important form of power, perhaps the most important form of power, is that exercised through the formation of the perceived interests of those dominated. He is right that B's subjective interests cannot be the baseline for detecting the exercise of power because they themselves may be the result of power. He is right that B's theoretically determined objective interests cannot be the baseline for detecting A's power because of the contestability of the theory involved and the potential for paternalistic tyranny. He is right that B's real interests are those which B would form when maximally independent of the power of others. But he is wrong to define power in terms of B's real interests. In fact, putting the argument of this paper together with Lukes's arguments, it is a mistake to define power in terms of B's interests at all, be they subjective, objective, or real.

My second point is not implausible, but requires some technical labor to make properly: you cannot properly identify power using a definition phrased in terms of intentions or wants, because intentions and wants may themselves be the creations of power. Who would define power in terms of wants? Recently Alvin Goldman[7] has,

[7] "Toward a Theory of Social Power," *Philosophical Studies,* XXIII (1972): 221–268.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

and done so in a very sophisticated theory of social power. Because his theory is so sophisticated, Goldman himself recognizes that the existence of radical power would require its rethinking. My allotted task will be to spell out in some detail how radical power defeats Goldman's enterprise, and thereby to suggest that the attempt to define power in terms of wants and intentions should be abandoned.

Goldman suggests that "[t]he core idea in the concept of power . . . is connected with getting what one wants" (*ibid.*, p. 222). But this simple formulation cannot be right because sometimes what one wants may be formed by what one gets. It is possible to adapt one's preferences to what one thinks will happen anyway.[8] Goldman, following Dahl, gives the example of the "chameleon" legislator "who always correctly predicts what the legislature is going to decide, and then forms a desire or preference to accord with this outcome" (*op. cit.*, p. 223). To avoid a definition that makes the chameleon powerful, Goldman concludes we must talk not only about getting what one actually wants, but also about getting what one would want, if one had other, hypothetical, desires. The chameleon gets what he wants, but not what he would have wanted had he been so rash as to want other than what he predicted would happen. "To explicate the concept of power, we must appeal therefore to subjunctive conditionals" (*op. cit.*, p. 223). Goldman suggests, then, the following analysis of power, where S is a person and E is an issue which may, or may not, have outcome e:

> S has power w.r.t. issue E if, and only if,
>
> (1) If S wanted outcome e, then e would occur, and
> (2) If S wanted outcome not-e, then not-e would occur (*op. cit.*, p. 225).

His analysis is later complicated by additional clauses that ensure that e or not-e occur because of some action performed by S, but these complications do not affect our argument. Goldman is interested in 'power to', not 'power over'. We can relate his analysis to our concern with A's power over B by supposing that S represents A, that the issue is B's behavior or welfare, and that the outcome is some particular change in B's welfare or behavior.

At the end of his paper, Goldman allows that, besides power over the welfare and behavior of others, there also exists,

> . . . the power to *persuade*, or more generally, power w.r.t. various psychological states of others, especially *desires*, *beliefs*, and *attitudes*.

[8] The phenomenon of adaptive preference formation is discussed in detail in Jon Elster, *Sour Grapes* (New York: Cambridge, 1985).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

> To the extent that a 'charismatic' leader has power w.r.t. the attitudes of his followers, and to the extent that the controller of communication media has power w.r.t. the desires and beliefs of the public, such men have power 'over' others (*op. cit.*, p. 263).

Goldman, however, sees immediately that the power to persuade will cause difficulty for his theory:

> According to our analysis, we are interested in whether an outcome would occur *if* agent [A] wanted that outcome. But suppose agent [C] is in a position to ensure that [A] will *not* want this outcome. What does this imply about [A]'s power w.r.t. that issue? This question must be taken up in a more fully developed theory of power, but this is not the occasion for such a development (*op. cit.*, pp. 263/4).

I confess I did not immediately see the problem, so I shall share the results of my investigation. Note, of course, that my inability to solve it does not mean the problem cannot be fixed by Goldman,[9] or by others.

The problem concerns the evaluation of the counterfactuals in Goldman's definition of A having power w.r.t. issue E. To evaluate a counterfactual conditional, we have to consider the truth values of its antecedent and consequent not only in the actual world, but at relevantly similar possible worlds. If we considered only the actual world, all counterfactual conditionals would be vacuously true because of the falsity of their antecedents. We must consider not only how things are, but how they might have been. But we need not consider everything that might have been; we need consider only states of affairs that do not differ drastically from the actual situation. To evaluate Goldman's counterfactuals at some time t_1, what we do is "hold constant the *resources* that S has at t_1 [which] is to say, roughly, that we hold constant *the state of the world at* t_1—with the exception, of course, of S's own desire vis-à-vis E and whatever is nomologically implied by that desire (e.g., the state of his brain)" (*op. cit.*, p. 232).

Given a set of relevantly similar worlds against which to do an evaluation, a counterfactual conditional will be true in two cases[10]:

[9] Goldman does not give an analysis of the power to persuade in his subsequent article, "On the Measurement of Power," this JOURNAL, LXXI, 8 (May 2, 1974): 231–252. See his comments on p. 234.

[10] I am following, but simplifying, David Lewis's treatment in *Counterfactuals* (New York: Blackwell, 1973), p. 16. My principal simplification is to ignore his use of a hierarchy of sets of possible worlds differing in their degree of similarity to the actual world. I believe this complication is irrelevant to my argument. Lewis's work was not available to Goldman when he wrote the article under discussion, and Goldman instead refers to the work of Robert Stalnaker in "A Theory of Conditionals," first published in 1968, and reprinted in Ernest Sosa, ed., *Causation and*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(a) It is true interestingly, if at every one of the relevantly similar possible worlds at which the antecedent is true, the consequent is also true.

(b) It is true vacuously, if there are no relevantly similar worlds where the antecedent is true.

In using this definition, the interesting work involves ascertaining what it means for a possible world to be relevantly similar to the actual one.

In the case in question, where *C* has power over *A*'s wants, the relevantly similar possible worlds have to be worlds where this power of *C*'s is operative. We want to know whether *A* has power over issue *E* in situations where someone else, *C*, has power over *A*'s wants. It would not make sense to look at worlds where *C* did not have this power, for that would be to apply the definition to some other case entirely. Whatever else we may change in considering similar but different worlds, we must surely hold constant the most important fact we know about the situation, i.e., *C*'s power over *A*'s wants. For instance, suppose that *A* is a very rich person, with vast economic resources. When evaluating the counterfactuals in the definition of *A*'s power, it would not make sense to consider, as relevantly similar, possible worlds where *A* did not have these resources. It would not make sense to try to identify whether prime minister *A* gets what she wants by supposing that she is not prime minister. In our case, it would not make sense to determine *A*'s power by considering cases where *A*'s wants are not determined by *C*.

But there is a problem. Suppose, firstly, that *C* has the power to ensure that *A* is indifferent or neutral w.r.t. *E*, and that *C* exercises this power. In other words, neither does *A* want *e* nor does *A* want not-*e*. According to Goldman's definition, *A* has power w.r.t. issue *E* if, and only if,

(1) If *A* wanted outcome *e*, then *e* would occur.
(2) If *A* wanted outcome not-*e*, then not-*e* would occur.

To evaluate these counterfactuals, we have to consider relevantly similar worlds. *C*'s power over *A*'s wants is operative at all the worlds relevant to either conditional, that is, *A* does not want outcome *e*, and *A* does not want outcome not-*e*. At all the relevantly similar worlds, the antecedents of both conditionals are false. So, by the truth definition for counterfactuals, both conditionals are vacuously true. When *C*'s power ensures that *A* is indifferent to the outcome, there are no hypothetical cases in which *A* wants either outcome. If

---

Conditionals (New York: Oxford, 1975), pp. 165–179. Although Lewis's account is the more generally cited, I believe my argument can also be made in Stalnaker's theory, though not as easily.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the conditionals are true, then, by Goldman's definition, *A* has power w.r.t. issue *E*. This, surely, is the wrong result. It entails that *A* has power over any issue about the outcome of which *A* is rendered indifferent by some other agent. Suppose prime minister *C* can manipulate citizen *A* into being indifferent to the result of the next election. By the definition, *A* has power w.r.t. the result of the election no matter who wins. The definition identifies *A* as powerful when *A* is most powerless.[11]

There may be some way to extend Goldman's analysis to cover conditioned neutrality, but I do not see how. We might try adding a third clause that would rule out applying his definition to cases where *A* is made neutral by someone else. We might, for example, add,

(3) If A is rendered neutral w.r.t. both e and not-e by some other social actor, then . . . .

The problem is how to fill in the blank. If we fill it in with '*e* would occur', we catch some of the cases, those when not-*e* occurs. But the definition still attributes power to *A* in those cases where *e* either occurs anyway or occurs because of *A*'s indifference. The opposite problem arises if we fill the blank in with 'not-*e* would occur'.

With a clause to such effect, we might try to limit the definition to cases without anyone's power over wants. This is a defeatist sort of strategy, because a definition that can handle power w.r.t. psychological states as well as w.r.t. behavior and welfare is to be preferred to one that cannot. But worse, it makes the definition of power unhelpful in identifying power. If we add,

(3′) *A*'s wants are not formed by the power of some other social actor.

[11] Goldman's full definition of power complicates this argument slightly. His full definition is:

S has power, at t_1, w.r.t. issue *E* (at t_n) if and only if
(1) There is a sequence of basic act-types such that
  (a) if *S* wanted *e*, then he would perform these acts at appropriate times between t_1 and t_n, and
  (b) if *S* performed these acts at these appropriate times, then *e* would occur (at t_n);
(2) There is a sequence of basic act-types such that
  (a) if *S* wanted not-*e*, then he would perform these acts at appropriate times between t_1 and t_n, and
  (b) if *S* performed these acts at these appropriate times, then not-*e* would occur (at t_n) (*op. cit.*, p. 226).

There are now four antecedents to falsify in order to render *S* vacuously powerful. *C*'s power to make *S* indifferent falsifies the antecedents to 1a and 2a. To falsify the antecedents to 1b and 2b, either *C* must also exercise power to prevent *S* from doing the relevant basic acts, or *S*'s wants in each case must be necessary conditions for *S*'s doing these acts.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

then we have to be able to identify the absence of the power of C and others over A's wants. To identify the absence of C's power, for example, we must first define it in terms of C's getting what C wants. This definition itself must contain the appropriate version of (3′), and we are again left pursuing the identification of power through an endless chain of possible worlds.

We should also consider the case where C has the power to ensure that A does not want some outcome. (The case where C ensures that A does want it is exactly parallel.) A problem arises when C is able to ensure both what A wants and what A gets. Suppose that C is the controller of communication media and can ensure both that A does not want a socialist party elected, and that few others want it either. It turns out that, by the definition, worker A is the one with the power over the socialist defeat. At all the relevantly similar worlds, C's power is operative, and the antecedent of the first clause of the definition, that A wants a socialist victory, is false at all of them. So the first conditional is vacuously true. The antecedent of the second conditional, that A wants a socialist defeat, is true at all the relevantly similar worlds, as is the consequent, that a socialist defeat occurs. So the second conditional is true, and, contrary-to-fact, A has power over the election outcome.

In cases where C does not have power over the outcome as well as over A's wants, C can still turn A into a chameleon. If C exercises her power to ensure that A always wants to vote with the majority, then the second conditional is vacuously true so long as A's predictions are infallible. The first conditional is nonvacuously true and the chameleon is powerful.

There seems to be no way, then, of defining power as getting what one wants. Similar considerations, I believe, apply to definitions like Bertrand Russell's[12] pithy definition of power as "the production of intended effects" (*ibid.*, p. 35), and to definitions in terms of A's perceived or subjective interests. A definition in terms of A's real interests will not work if real interests are defined in the manner of Lukes. The above considerations would not apply to a definition in terms of objective interests, if these existed and could be identified, because objective interests are presumably independent of power.[13]

[12] *Power: A New Social Analysis* (New York: Norton, 1938).

[13] It is worth noting that a similar sort of problem arises for definitions of freedom in terms of want satisfaction. It would not be wise to define A's freedom to accomplish x in terms of the material conditional that, if A wants to accomplish x, A will accomplish x, because the conditional is true, and A is free, whenever A does not want to accomplish x. If C puts A in jail, and A does not, for whatever reason, want to leave, A is thus free. To avoid this consequence, we could try the counterfactual conditional that, if A wanted to accomplish x, then A would accomplish x. If we evaluate this counterfactual, holding constant the condition that C has put A in

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

My third point is this: you cannot define the concept of power in terms of one agent's being responsible for affecting another because an agent's responsibility entails his independence from the power of others. The term to be defined, 'power', will reoccur in its own definition. My final moral is simple: power must be defined prior to locating responsibility for its exercise.

Who would define power in terms of responsibility, and why? The political theorist, William Connolly[14] does, and he does so, again, for laudable reasons. Like Lukes, Connolly is interested in locating responsibility for the effects of power. He believes that in this way the concept of power can be a tool for the criticism of society:

> . . . to acknowledge power over others is to implicate oneself in responsibility for certain events and to put oneself in a position where *justification* for the limits placed on others is expected. To attribute power to another, then, is not simply to describe his role in some perfectly neutral sense, but is more like *accusing* him of something, which is then denied or justified (*ibid.*, p. 97).

Responsibility comes in degrees; A is responsible to the extent that someone else does not exert power over A. This extent ranges from the maximally powerful use of force to the effect of rational persuasion which does not involve power at all:

> Because of the ways in which manipulation, coercion, anticipatory surrender, deterrence, force and conditioning impinge on the choices of recipients, we shift, in some degree, responsibility in these cases away from the recipient and toward the person initiating these pressures. Only in the case of persuasion, which is not a form of power, does such a shift not occur . . . (*ibid.*, p. 95).

Because of its "close conceptual connection" to power, Connolly builds responsibility right into his definition of power, which he summarizes as:

> Briefly stated, A exercises power over B when he is responsible for some x that increases the costs, risks, or difficulties to B in promoting B's desires or in recognizing or promoting B's interests or obligations (*ibid.*, p. 102).

---

jail, we get the right result. But, as before, suppose that C makes A not want to accomplish x, and hold this condition constant while evaluating the counterfactual. Now C makes A free by forming A's wants. We can do better if we attempt a similar definition for unfreedom. The relevant counterfactual—that, if A wanted to accomplish x, A would fail—is made vacuously true by C's power to make A not want x, and thus C renders A unfree.

[14] *The Terms of Political Discourse* (Boston: Heath, 1974).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

In order to be able to identify when *A* is exercising power in doing *x*, we must first ascertain *A*'s responsibility for *x*. Connolly gives a list of conditions for the ascription of responsibility. To be fair, he proposes these conditions not as a criterion but as a paradigm of when responsibility is correctly ascribed. The list includes:

(a) *A* introducing *x* *intentionally*.
(b) *A* introducing *x* unintentionally, but *foreseeing* the consequences for *B*.
(c) *A* having a *disposition* for introducing *x* unless *B* acts to limit *B*'s interest, and *B* believing this.
(d) *A* failing to prevent *x* though *understanding* the consequences of *x* for *B*.[15]

As I have emphasized, in each of these conditions of responsibility, and in any others that could reasonably be proposed, *A*'s mental attitude toward *x* figures prominently. What *A* intends, foresees, understands, or is disposed to do are important to the degree of responsibility which *A* bears for both *x* and its effect on *B*. One reason for using a notion of moral or legal responsibility is to broaden the definition of power to cases, such as the careless use of power, which are missed by an account in terms of *A*'s wants or intentions.

Responsibility comes in degrees, ranging from full through partial to none, and this is relevant to identifying power. Consider the following case: a dictator sends a soldier to arrest a dissident. The dictator makes it clear that, if the soldier does not do as ordered, he will face a firing squad. There is a sense in which the soldier is responsible for arresting the dissident, because he could have done otherwise. He could have chosen not to arrest her and sought to escape the firing squad as best he could. If he sympathized with her cause, he might have deliberated on the pros and cons, and chosen, as a rational agent, what to do. But there is another sense in which he is not responsible. The power of the dictator is so great, the risk of the soldier's capture so high, and the sanction so extreme that he does not have much choice in the matter. The soldier is both responsible and not responsible, powerful and powerless, in arresting the dissident.

We can easily resolve this conundrum by saying that the soldier is partially but not fully responsible for arresting the dissident. The dictator bears full responsibility for ordering the arrest, since we do

[15] This is my summary of a list of conditions on p. 102 of Connolly, *op. cit.* The emphasized words are all his own with the exception of 'disposition', which is my summary of having 'a history that reasonably leads *B* to believe that an *x* might well be introduced'.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

not imagine anyone exerting power over him. The soldier's responsibility is diminished by the coercive power of the dictator. The notion of partial responsibility is useful to Connolly's analysis; a decision that one agent is only partially responsible for an act points beyond that agent to another locus of power. It would be a very inadequate critique of the dictator, soldier, and dissident society which ceased its analysis with the soldier's responsibility.

The important conclusion to draw from this story, however, is another conceptual connection between responsibility and power: *A*'s degree of responsibility can be diminished when someone else exerts power over *A*. In fact, there are circumstances in which *A* would be said to have no responsibility at all, and no power. Suppose the dictator sends two secret police agents with the soldier. This time the task is to murder the dissident. The police put the pistol in the soldier's hand, point it, and push his trigger finger. The soldier has no responsibility, and no power over the dissident, because the police have complete power over him.[16]

I am inclined to say the following about these two cases: in both cases, the dictator has full responsibility, and thus power; in the arrest case, the soldier has partial responsibility and does exercise power, or at least some power; and in the murder case, the soldier has no responsibility, and no power, because he is completely in the power of others.

The upshot of this connection between responsibility and power is that *A* is responsible only if there does not exist some other person or persons who have complete power over *A*. A necessary condition for *A*'s responsibility is that *A* not be in the complete power of another. We cannot identify responsibility without first being able to identify the absence of power, and we cannot identify the complete absence of power without already having a definition of power. So a definition of the concept of power framed in terms of *A*'s responsibility will be a definition of power in terms of power.

It might be thought that Connolly has anticipated this objection. I have, perhaps, not adequately represented his position; in the first part of his four-part criterion for assigning responsibility, he does not say just that *A* must introduce *x* intentionally, but that *A* must

[16] Harry Frankfurt has argued that not only total physical constraint, but also an extreme coercive threat, can remove moral responsibility. He writes: "If the victim's desire or motive to avoid the penalty with which he is threatened is—or is taken by him to be—so powerful that he cannot prevent it from leading him to submit to the threat, then he really has no alternative other than to submit. He cannot effectively choose to do otherwise. It is only then that it *may* be proper to regard him as bearing no moral responsibility for his submission"; "Coercion and Moral Responsibility," in Ted Honderich, ed., *Essays on Freedom of Action* (Boston: Routledge & Kegan Paul, 1973), p. 77.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

"introduce *x* as an intentional act, where *A* could have done otherwise" (*op. cit.*, p. 102). The second part of the clause, that *A* could have done otherwise, is either explicit or implicit in the other parts of his criterion. Andrew Reeve[17] comments that the second part of the condition,

> . . . is presumably required by the feature to which Frankfurt draws attention: coercion exists where the victim has no choice, and no moral responsibility. It might happen that the victim was A, introducing x as an intentional act, but himself under the power of some other individual. In such circumstances, A's intention, but lack of choice, does not make him responsible . . . (*ibid.*, pp. 82/3).

But anticipating and meeting the problem of power diminishing responsibility are two different things. Adding the clause, '*A* could have done otherwise', just serves to make the circularity of the definition explicit. If Reeve's interpretation of the clause is correct, it could be replaced by the condition that *A* is (1) *not* in the complete power of some other social actor, and (2) *not* prevented by natural obstacles from doing otherwise. But the first part of this new condition is enough to show that, once again, in order to identify the presence of power through the ascription of responsibility, we must first be able to identify its absence.

It might also be thought that situations in which someone is in the *complete* power of another are extremely obvious, so obvious in fact that they do not require a definition of power to identify them. The physical control exerted by the police over the soldier in the murder case is perhaps such a situation. If we could reduce being in the complete power of another to being under the total physical control of another, we could, without circularity, define power and responsibility in terms of the absence of total physical control.

But forms of complete power may be more prevalent and less obvious than the objection assumes. Suppose that Lukes is right about the existence of radical power, power exercised over a person "by influencing, shaping or determining his very wants." Recall that, for Connolly, any ascription of responsibility to *A*, no matter how loose, must involve *A* having certain mental attitudes. *A* must intend, foresee, be disposed to, or understand something. These mental attitudes are potentially the sites of radical power, open to forms of power in which the compliance of others is secured "by controlling their thoughts and desires."

Radical power can be both complete and self-concealing, totally controlling those it affects yet totally invisible to them. Suppose the

[17] "Power Without Responsibility," *Political Studies*, xxx (1982).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

dictator sends the soldier to interrogate the dissident, only this time the dictator controls the soldier through mechanisms of propaganda and misinformation which legitimize the dictator's authority. Given the censored information and distorted values to which he is exposed, the soldier makes a rational decision to interrogate the dissident, a decision that any reasonable person in a similar situation would make.[18] Is the soldier responsible for the interrogation? The dictator's power over the soldier may be as absolute as his physical control in the murder case, yet it may be invisible to all but the most sensitive of social criticism. In order to hold the soldier responsible, we must first be able to determine conclusively the absence of subtle forms of power.

Connolly, quite laudably, wants to use the analysis of power as a tool for social change, as a way of assigning responsibility to social actors. But he fails to see that power must be identified prior to the ascription of responsibility in order to avoid a regress.

My final conclusion is this: generalizing from the failure of the above three definitions, I would suggest that human agency is not the place to look for an adequate conception of power. If radical power exists, it will distort human agency by manipulating the beliefs and desires that ground rational decision making. A definition of social power in terms of human agency will be made difficult by the absence of a concept of a pure and unadulterated agent on which to base it. The existence of the power of one person over another is a social fact, not a material one; people do not have power because of some intrinsic property of themselves, but because of the social relations in which they are embedded. They have power, or are subject to it, because of their position in a social structure or because of their role in a social system. A fruitful way of identifying social power would disclose power in all its guises, and point beyond itself to the real source of that power.

ANDREW KERNOHAN

Dalhousie University

[18] On the question of when persuasion becomes an exercise of power, see Stanley Benn, "Freedom and Persuasion," *The Australasian Journal of Philosophy*, XLV (1967): 259–275.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

COMMENTS AND CRITICISM

INTENSIONALITY AND PERCEPTION: A REPLY TO ROSENBERG

PERCEPTUAL states purport to inform us of situations external to ourselves, and they do so by means of their descriptive content. Let us say that I am subject to certain visual sensations. By the intrinsic character of these sensations, I am able to identify them as purporting to inform me that I am in the presence of something red. In this case, *red* is the descriptive content of my visual state.

It is a commonplace of epistemology that perception gives me a good (but not conclusive) reason for supposing that I am in the presence of something that possesses the property that constitutes the content. But the commonplace raises puzzling questions. What makes it the case that a state with this intrinsic character informs me of the presence of a red thing, rather than of the presence of a blue thing? In other words, how do the intrinsic characters of states get matched up with the messages they carry? This is the problem of meaning for perceptual states. Second, given that a state of this type is matched up with the message that something red is in the vicinity, what makes the message it delivers credible?

In "Biological Functions and Perceptual Content,"[1] I undertook to analyze what it means to endow a perceptual state with content, and I hoped to do so in such a way as to make the credibility of perception intelligible. I took it as given that any analysis of this sort would have to conform to two constraints. First, it must treat perception as having meaning independently of any conventional significance that we attach to it. Thus, the problem of meaning for perception cannot be solved in the same way as the problem of meaning for language. Secondly, it must treat perception as a natural phenomenon. *Pacé* Descartes, the credibility of perception cannot be grounded in the goodness of the creator, for to treat perception as created is antithetical to treating it as a natural phenomenon.

In a recently published comment, Alexander Rosenberg[2] argues that my analysis fails for two reasons. First, he says, it is unsuccessful in accounting for a logical feature of verbs of perception—their intensionality. Second, he suspects that I mask this defect by appealing to a concept that contains in itself an unanalyzed reference to

[1] This JOURNAL, LXXXV, 1 (January 1988): 5–27; hereafter BFPC.
[2] "Perceptual Presentations and Biological Functions," this JOURNAL, LXXXVI, 1 (January 1989): 38–44.

0022-362X/89/8612/727–733 © 1989 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

content, a move that Rosenberg characterizes as being "in the worst traditions of homuncularism" (41). My purpose in this note is to defend myself against these criticisms.

The analysis that Rosenberg objects to is the following:

> A perceptual state is a state that has the [biological] function of *detecting* the presence of things of a certain type, where things of that type are sometimes present, sometimes absent.
>
> A perceptual state with internal characteristic S [has descriptive content] F if and only if its function is to *detect* the presence of a thing with property F (BFPC, 20, with inessential changes).

Rosenberg thinks that detecting might itself be an intentional notion, and says rightly that, if this is so, my analysis is worthless. I shall address this worry first, since to do so will allow me to give a brief self-contained exposition of my position concerning perceptual content.

Let me start with a definition. I shall say that a state with characteristic S indicates the presence of a thing with property F (or more briefly that S indicates F) if such states occur only when caused by a thing with property F.[3]

One simple causal theory of perceptual content states that, if a state indicates F, then its content is F; for if this were so one could appreciate why perception is a reliable indicator of F.[4] But this account will not do for three reasons. First, not every indicator state has content. It might be, for instance, that the chemical composition of my blood indicates that I have had more than one alcoholic drink, but this does not mean that the chemical composition of my blood has my inebriation level as its content. Second, it is not strictly true that perceptual states indicate their content. We are prone to misperception, after all, and this shows that a perceptual state can occur even when it is not caused by a thing with the property that constitutes its descriptive content. Third, not everything that a perceptual state indicates is included in its content. For example, one cannot see something as blue without a certain region of the visual cortex being excited, or without one's eyes being open, or without there being sufficient light. The perception of something as blue indicates all of these things, but none is part of its content.

These defects of the indicator view can be rectified, I claim, by

[3] Indication might be thought to be a matter of degree. If so, the following definition is more appropriate: S indicates F if S is caused by F with sufficient frequency to make it reasonable to infer the presence of F from the occurrence of S.

[4] Of course, this does not wholly address the problem, which was not why perception is reliable, but why it is credible—that is, why it has epistemic force.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

invoking biological functions. Suppose that the content of a perceptual state is F not when it does in fact indicate F, but when its *biological function* is to indicate F. Then the above-mentioned difficulties disappear. First, though the chemical composition of my blood does in fact indicate how many drinks I have had, it is not its biological function to do so. This is why the composition of my blood is not a perception of the amount of alcohol I have consumed. Second, even if a thing has F as its function, it might not always do F—e.g., my windshield wipers have the function of wiping the water off my windshield, but at times they are unable to do so. Misperception, then, does show that our perceptual states are not perfect indicators; but even so it might be their function to indicate. Third, the fact that perceptual states indicate all sorts of things that are not contained in their content is no longer bothersome, for they might not have the biological function of indicating everything that they do in fact indicate. The visual system, for example, was not "designed" to ferret out its own states, though in fact these states are deducible from visual perceptions. Finally, the biological-function account helps us understand why perception is credible: because it is a product of evolution.[5]

But the biological-function view is incomplete. The trouble is that all sorts of states have the biological function of indicating the occurrence of various conditions without its being true that they are perceptions of these conditions. The example offered in my article was that of states of the bacterium *e. coli*—it appears that the bacterium has states, the function of which is to covary with states of the environment and to control its reproductive rate accordingly. Another example is that of states of such nonperceptual organs as the mammalian immune system.[6] It is the function of this system to register the presence of intruders and to destroy these when they are found. If what interests us is perception as such—the epistemic faculty that is relevant to our formation and assessment of beliefs—we ought not to be satisfied with a definition that applies also to states of bacteria (which presumably lack beliefs), or to states of human systems like the immune system (which are irrelevant to belief).

[5] I do not suggest that this is a reason for supposing that perception is perfectly credible. Reliability is one selectional constraint on the evolution of perception, but only one. Speed, energetic efficiency, and the nature of the evolutionary predecessors of perceptual organs are other constraints. It is by reference to these other constraints and to the inherent difficulties in detecting the occurrence of external states of affairs that perceptual error is accommodated in my account.

[6] See Ed Levy and my, "Teleology, Error, and the Human Immune System," this JOURNAL, LXXXI, 7 (July 1984): 351–372.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

A simple and direct way to address this difficulty is to make a distinction based on what functional indicator states are used for. Control states like those in *e. coli* and the immune system are in direct contact with the actions performed by those entities: in contact with a changing environment, they automatically adjust the behavior of the system they serve so that it is appropriate to the environment. Perceptual systems must surely have evolved from such automatic control systems,[7] but they do not serve directly to initiate action; rather, they interact with memory, reason, desire, and so on—in short, with the faculty of judgment. If action results, it comes out of judgment, not directly out of perception (BFPC, 20).

It was to this epistemic character of perception that I wanted to draw attention by the use of the term 'detection'. The difference between control states ("quasi-perceptual states" as I called them) and perceptual states, then, is not a matter of how they get their content (both get their content from their biological function of indicating the presence of a thing with a certain property), but of how they are used within the organism. Quasi-perceptual states serve to initiate action appropriate to the situation they are supposed to indicate, perceptions proper serve to provide input to the epistemic faculty.[8] But both have the function of indicating the occurrence of an external situation of a certain sort, and the match between these states and the messages they carry can be made without the use of unanalyzed intentional notions. With respect to their intentionality, then, the two sorts of states are similar. There is no substance in Rosenberg's suspicion that my use of the notion of detection was "in the worst traditions of homuncularism."

It must be acknowledged that, unfortunately, this statement leaves the precise nature of the interaction between perception and the epistemic faculty undefined. The reason is that this interaction must be intentional, at least in the first analysis, in some such way as this: the perception of something as F has the function of causing the

[7] The strongest evidence that they have so evolved is the existence of automatic control systems physiologically similar to our organs of perceptions. Vision in the frog and in many insects is largely an automatic control system: frogs automatically flick their tongues when they see flying insects, and flies automatically adjust their landing mechanisms according to the sensed distance of the landing pad. By my classification these would not be perceptual mechanisms as such, any more than the mammalian immune system, because like the latter they have no epistemic role to play, but simply initiate behavior appropriate to external circumstances.

[8] In "Biosemantics," this JOURNAL, LXXXVI, 6 (June 1989): 281–297, Ruth Millikan criticizes me for not taking the "consumer" of a perceptual representation into account, as well as its producer. Although the distinction was not clear enough in BFPC, or in my own mind until I talked to her, I believe that it was implicit there that the distinction between quasi-perceptual states and perceptual states lies in the fact that the former feed into motor control and the latter into the epistemic faculties.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

perceiver to believe that the thing is *F*, in the absence of reasons for believing otherwise. Until one is able to give an account of what it is to endow belief states with content, it will not be possible to eliminate intentionality from this sort of first analysis. It is beyond the scope both of BFPC and of the present paper to offer such an account. I can say, however, that I do not think that there is any straight-forward generalization of the biological function account of perceptual content to cover the content of belief.[9] This is, firstly, because it does not seem plausible to assign belief states content on the basis of their intrinsic features. There is no good reason to think that there is a materialistic type-type reduction of intentional content for belief,[10] and it is an essential feature of the attribution of biological functions that there should be a physical type of thing to which these functions are attributed. Secondly, it is implausible to suppose that individual beliefs have functions. Does the state that embodies my belief that $E = mc^2$, for instance, have the function of indicating the truth of that proposition? Obviously not. Does the belief that it is time to eat have a function? Not obviously. For these reasons, it seems to me that the content of belief states will have to be assigned derivatively. First we shall have to assign functions to inferential routines in the human brain—modus ponens, various inductive schemata, and so on.[11] Then the products of these inferential routines—beliefs—can be assigned content.

Let us move on now to Rosenberg's other criticism that my account, in common with all naturalistic accounts, fails to account for the intensionality of perceptual contexts. I shall start by detailing the ways in which, as I understand it, perceptual verbs generate intensional contexts.

Distinguish first between the *object* of a perceptual state and its *descriptive content*. Suppose I am looking at a white billiard ball. The thing I see, the object of my visual state, is the billiard ball itself, under any description. If it happens to be Rosenberg's only cue ball, then what I see is Rosenberg's only cue ball, whether I know it or not.[12] With respect to the object of perception, then, the verb 'sees'

[9] In fact, Millikan does seem to give an account of belief contents which is broadly similar to my account of perceptual states: see her *Language, Thought and Other Biological Categories: New Foundations for Realism* (Cambridge: MIT, 1984), esp. chs. 5 and 6, and now "Biosemantics."

[10] There is a vast amount of neurophysiological data that indicates that type-type reduction is plausible for perceptual states; for a review, see David Hubel, *Eye, Brain and Vision* (New York: Freeman, 1988).

[11] This assumes, of course, that these can be reduced type for type.

[12] The point is made by G. J. Warnock in "Seeing," *Proceedings of the Aristotelian Society*, LV (1954–55); reprinted in Robert J. Swartz, ed., *Perceiving, Sensing, and Knowing* (Garden City, NJ: Doubleday, 1965), pp. 49–65.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

accepts coreferential terms as object with no change in truth value. In this respect, these contexts are extensional.

The fact that this particular object is what I see is not, however, written directly into the perceptual state. What is available there is simply that the object that I am looking at, whatever its identity, has certain properties—in this case, that it is white and spherical—and that it is related to me in certain ways. This is what I call the descriptive content of the visual state; it is what I see the object as, to use another common way of talking about these things.

I see the billiard ball as white and spherical. Now, suppose the predicate 'white and spherical' is coextensive with the predicate 'a billiard ball in this room'. Even if this is so, it does not follow from my seeing the ball as white and spherical that I see it as a billiard ball in this room. This is true for a variety of reasons, but one crucially relevant point is that not all properties are accessible to me by sight as such. There are certain properties that are (as Plato and Aristotle used to put it) proper to sight, and whiteness and sphericity are among these.[13] Being a billiard ball is not such a property. It is therefore not something "written into" my visual state that the thing that I am looking at is a billiard ball, and this holds whether or not I know that it is. Thus, even though the terms are coextensive, I see something as white and spherical, but not as a billiard ball in this room: the contexts generated by 'sees as' are not therefore truth-preserving under the substitution of co-extensive predicates. In this sense, they are intensional.

Are these logical features of the verb 'to see' preserved by the account of perceptual content given above? On the face of it, my account does preserve these features. So does every other broadly similar causal account—the appeal to functions is not crucial to this point.

First, note that I follow John Searle in claiming that the object of a perceptual state is whatever satisfies the following description: the individual that has the property *F* and is causing *this* state.[14] Since the object so specified is an individual as such and not an individual "under a description," the object of perception will be so, however it

[13] Philosophers have said that it is just the two-dimensional projection of sphericity that is proper to sight, on the grounds (a) that it is the two-dimensional projection that affects the retina, and (b) that errors are possible with respect to sphericity. It is a virtue of the computational theories of perception that I was operating with that they make (a) irrelevant to the question, and a virtue of my account that it makes (b) irrelevant (since it allows error with respect to the proper objects of perception).

[14] In fact, there are some problems with this specification of the object of a perceptual state, but they need not concern us here.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

is described. Thus, the tendency of the verb 'see' to accept substitutions of coreferential terms in the direct-object position is preserved. So this aspect of the logic of perception is well-served by the logic of causation.

Next, let us look at descriptive content. A causal account of descriptive content says that it is relevant to a perceptual state having content *F* that is caused by things that are *F*. Contrary to what Rosenberg says, the logic of causation makes it illegitimate to substitute coextensive predicates for *F* here. That my blood circulates is due to the fact that I am a creature with a heart. Having a heart is a property coextensive (let us allow) with having a kidney. Nevertheless, that my blood circulates is not caused by my having a kidney. So causal contexts do not permit the substitution of coextensive terms.[15] It is therefore open to the causal theorist to say that my perceptual state is caused by something's being white and spherical but not by its being Rosenberg's only cue ball—even though the terms are co-extensive. Again, there seems to be no problem as far as the logical properties of perception are concerned.

What applies to causal contexts generally applies even more to the contexts generated by 'biological function'.[16] Suppose that it is the biological function of this visual state to detect the presence of a white thing. 'White' is coextensive with 'has the same color as alabaster'. It does not follow that the function of this type of state is to detect the presence of things that have the same color as alabaster. I analyze biological functions in terms of evolutionary causes, and Rosenberg argues that, since all causal contexts are extensional with respect to substitutions of coextensive terms, my account is inadequate with respect to the intensionality of perception. It is his premise that is at fault: there is no reason to think that causal contexts are extensional in this respect.

MOHAN MATTHEN

University of Alberta

[15] One would have thought that Rosenberg would be receptive to the distinction between intensionality with respect to coextensive predicates and intensionality with respect to coreferential terms. Ten years ago, he and R. M. Martin wrote: "The admission that causal statements fail the test of extensionality in the case of substitutivity of co-extensive predicates into sentence arguments is not very damaging, since it is tantamount to the admission that causal statements are not truth-functional, and this is something we know already"; "The Extensionality of Causal Contexts," in *Midwest Studies in Philosophy*, IV (1979): 401–408. The quoted sentence occurs on p. 406, and is followed by an obscure demand that we should modify our understanding of co-extensionality.

[16] Elliott Sober's distinction between selection *of* objects and selection *for* properties that these objects possess is very relevant here. See his *The Nature of Selection* (Cambridge: MIT, 1984), pp. 97–102.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

NOTES AND NEWS

The Fifth Annual Conference on Computers and Philosophy will be held at Stanford University on August 9–11, 1990. The conference is sponsored by the American Association of Philosophy Teachers, the Committee on Computer Use in Philosophy of the APA, and the Department of Philosophy, the Symbolic Systems Program, CSLI, and IMSSS at Stanford University. The Program Committee consists of John Etchemendy (chair), Jon Barwise, and Tryg Ager. Presentations on all aspects of the relationship between philosophy and computers are invited, including ethical issues surrounding computer use, the philosophical foundations of computation, the use of computers in teaching philosophy and logic, and issues of program verification. Five-page abstracts in triplicate may be sent, by March 15, 1990, to John Etchemendy, Center for the Study of Lang. and Info., Ventura Hall, Stanford Univ., Stanford, CA 94305. Proposals for the site of the 1991 meeting may be submitted, by January 15, 1990, to Robert J. Cavalier, CDEC, Carnegie Mellon Univ., Pittsburgh, PA 15213.

The Association for Philosophy of Education is pleased to announce a call for papers for its December 1990 meeting, to be held in conjunction with the 1990 Eastern Division APA Meeting. The deadline for papers, which should be twenty minutes reading time, is March 1, 1990. Papers may be submitted in triplicate to Sophie Haroutunian-Goedon, Dept. of Education, University of Chicago, 5835 S. Kimbark Ave., Chicago, IL 60637.

National Endowment for the Humanities: Since its inception seventeen years ago, the Summer Seminars for College Teachers program has provided opportunities for over 14,000 college faculty members and unaffiliated humanities scholars to study in a collegial atmosphere under the direction of distinguished scholars in their own or a related discipline. NEH seminars are held at major research centers, where participants can pursue their own research and writing projects. This summer, 51 seminars will be offered on such topics as: the Philosophical Uses of Historical Traditions; Ancient Philosophy of Law; Consequentialist Theories of Morality; Heidegger and Davidson: Critics of Cartesianism; the Nature and Value of Autonomy; the Philosophical Foundations of the Social Sciences; the Philosophy of Socrates; Religion and Society in Hegel, Marx, and Kierkegaard; the Spectrum of Responsibility; and Supervenience and its Philosophical Applications. Applications are invited from prospective participants. Also, plans are underway for the selection of 1991 Summer Seminar Directors, for which the Endowment encourages applications from scholars with distinguished teaching and publishing records. Proposed topics should focus on enduring issues or current scholarship in the humanities. The application deadline from participants and prospective directors is March 1, 1990. Further information may be obtained by writing SSCT, Room 316, Division of Fellowships and Seminars, NEH, 1100 Pennsylvania Ave., Washington, DC 20506.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

U.S. Postal Service

STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION

Required by 39 U.S.C. 3685)

1A. TITLE OF PUBLICATION	1B. PUBLICATION NO.	2. DATE OF FILING
The Journal of Philosophy	0 0 2 2 3 6 2 X	8/15/89
3. FREQUENCY OF ISSUE	**3A. NO. OF ISSUES PUBLISHED ANNUALLY**	**3B. ANNUAL SUBSCRIPTION PRICE**
Monthly	12	$20.00 Student $30.00 Indiv $50.00 I

4. COMPLETE MAILING ADDRESS OF KNOWN OFFICE OF PUBLICATION *(Street, City, County, State and ZIP+4 Code) (Not printers)*

709 Philosophy Hall, Columbia University, New York, N.Y. 10027

5. COMPLETE MAILING ADDRESS OF THE HEADQUARTERS OF GENERAL BUSINESS OFFICES OF THE PUBLISHER *(Not printer)*

709 Philosophy Hall, Columbia University, New York, N.Y. 10027

6. FULL NAMES AND COMPLETE MAILING ADDRESS OF PUBLISHER, EDITOR, AND MANAGING EDITOR *(This item MUST NOT be blank)*

PUBLISHER *(Name and Complete Mailing Address)*

The Journal of Philosophy, Inc.
709 Philosophy Hall, Columbia University, New York, N.Y. 10027

EDITOR *(Name and Complete Mailing Address)*

Arthur Danto, 709 Philosophy Hall, Columbia University, New York, NY 10027

MANAGING EDITOR *(Name and Complete Mailing Address)*

Michael Kelly, 704 Philosophy Hall, Columbia University, New York, N.Y. 10027

7. OWNER *(If owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1 percent or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual must be given. If the publication is published by a nonprofit organization, its name and address must be stated.) (Item must be completed.)*

FULL NAME	COMPLETE MAILING ADDRESS
The Journal of Philosophy, Inc.	709 Philosophy Hall Columbia University New York, N.Y. 10027

8. KNOWN BONDHOLDERS, MORTGAGEES, AND OTHER SECURITY HOLDERS OWNING OR HOLDING 1 PERCENT OR MORE OF TOTAL AMOUNT OF BONDS, MORTGAGES OR OTHER SECURITIES *(If there are none, so state)*

FULL NAME	COMPLETE MAILING ADDRESS
NONE	

9. FOR COMPLETION BY NONPROFIT ORGANIZATIONS AUTHORIZED TO MAIL AT SPECIAL RATES *(Section 423.12 DMM only)*
The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes *(Check one)*

(1) [X] HAS NOT CHANGED DURING PRECEDING 12 MONTHS
(2) [ ] HAS CHANGED DURING PRECEDING 12 MONTHS

(If changed, publisher must submit explanation of change with this statement.)

10. EXTENT AND NATURE OF CIRCULATION *(See instructions on reverse side)*	AVERAGE NO. COPIES EACH ISSUE DURING PRECEDING 12 MONTHS	ACTUAL NO. COPIES OF SINGLE ISSUE PUBLISHED NEAREST TO FILING DATE
A. TOTAL NO. COPIES *(Net Press Run)*	4845	4737
B. PAID AND/OR REQUESTED CIRCULATION 1. Sales through dealers and carriers, street vendors and counter sales	----	----
2. Mail Subscription *(Paid and/or requested)*	3961	3908
C. TOTAL PAID AND/OR REQUESTED CIRCULATION *(Sum of 10B1 and 10B2)*	3961	3908
D. FREE DISTRIBUTION BY MAIL, CARRIER OR OTHER MEANS SAMPLES, COMPLIMENTARY, AND OTHER FREE COPIES	174	178
E. TOTAL DISTRIBUTION *(Sum of C and D)*	4135	4086
F. COPIES NOT DISTRIBUTED 1. Office use, left over, unaccounted, spoiled after printing	710	651
2. Return from News Agents	----	----
G. TOTAL *(Sum of E, F1 and 2—should equal net press run shown in A)*	4845	4737

11. I certify that the statements made by me above are correct and complete

SIGNATURE AND TITLE OF EDITOR, PUBLISHER, BUSINESS MANAGER, OR OWNER

Michael Kelly Manager, Editor

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

INDEX TO VOLUME LXXXVI

NAMES OF CONTRIBUTORS ARE PRINTED IN SMALL CAPITALS

Abortion (Why — Is Immoral). DON MARQUIS. 183 (4).

ACKERMAN, BRUCE: Why Dialogue? 5 (1).

Adam Smith on Virtue and Self-interest [abstract]. CHARLES L. GRISWOLD, JR. 681 (11).

Agency (Social Power and Human —). ANDREW KERNOHAN. 712 (12).

ALANEN, LILLI: Review of John Cottingham: *Descartes*. 44 (1).

ALLISON, HENRY E. Review of Paul Guyer: *Kant and the Claims of Knowledge*. 214 (4).

American Philosophical Association (APA), Eastern Division, 86th Annual Meeting. Symposium Papers: 505 (10), 601 (11); Program: 567 (10).

Analytical Marxism. APA Symposium. 659 (11).

Applying Kripke's Theory of Truth. VANN MCGEE. 530 (10).

Backward Induction Paradox (The —). PHILLIP PETTIT and ROBERT SUGDEN. 169 (4).

BAKER, JUDITH: The Metaphysical Contruction of Reality. 505 (10).

Bayesian (On the Shared Preferences of Two — Decision Makers). TEDDY SEIDENFELD, JOSEPH B. KADANE, and MARK SCHERVISH. 225 (5).

Beliefs, Puzzling (The Prince and the Phone Booth: Reporting —). MARK CRIMMINS and JOHN PERRY. 685 (12).

Benhabib, Seyla: *Critique, Norm and Uptopia: A Study of the Foundations of Critical Theory*. Review by ELLIOTT L. JURIST. 203 (4).

BILGRAMI, AKEEL: Realism without Internalism: A Critique of Searle on Intentionality. 57 (2).

Biosemantics. RUTH GARRETT MILLIKAN. 281 (6).

Bodily Inscriptions (Foucault and the Paradox of —). JUDITH BUTLER. 601 (11).

Body (Culture or Nature? The Function of the Term '—' in the Work of Michel Foucault). LADELLE MCWHORTER. 608 (11).

Brute Experience. PETER CARRUTHERS. 258 (5).

BURIAN, RICHARD M.: Review of Philip Kitcher: *Vaulting Ambition: Sociobiology and the Quest for Human Nature*. 385 (7).

BUTLER, JUDITH: Foucault and the Paradox of Bodily Inscriptions. 601 (11).

CAMPBELL, JOHN: Review of Gareth Evans: *Collected Papers*. 156 (3).

CARRUTHERS, PETER: Brute Experience. 258 (5).

Cartwright and the Lying Laws of Physics. RONALD LAYMON. 353 (7).

Computation and Consciousness. TIM MAUDLIN. 407 (8).

Consciousness (Computation and —). TIM MAUDLIN. 407 (8).

Consequentialism (A Paradoxical Departure from —). GEORGE W. HARRIS. 90 (2).

Copyright (Statement of — Policy). 56 (1).

COTTINGHAM, JOHN. *Descartes*. Review by LILLI ALANEN. 44 (1).

CRIMMINS, MARK, and JOHN PERRY: The Prince and the Phone Booth: Reporting Puzzling Beliefs. 685 (12).

Culture of Nature? The Function of the Term 'Body' in the Work of Michel Foucault. LADELLE MCWHORTER. 608 (11).

DARWALL, STEPHEN L. Review of W. D. Falk: *Ought, Reasons, and Morality*. 208 (4).

Degrees (On —). RAYME E. ENGEL. 23 (1).

Dialogue (Why —?). BRUCE ACKERMAN. 5 (1).

Domain Extension and the Philosophy of Mathematics. KENNETH MANDERS. 553 (10).

DRIVER, JULIA: The Virtues of Ignorance. 373 (7).

Egoism and Morality. APA Symposium. 542 (10).

Egoism and Morality. LESLIE MULHOLLAND. 542 (10).

Egoism (Doing Justice to—) [abstract]. LAURENCE THOMAS. 551 (10).

ENGEL, RAYME E.: On Degrees. 23 (1).

Epistemic (The Maladroitness of — TIT FOR TAT). JOHN WOODS. 324 (6).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Errata. 112 (2).
Evans, Gareth. *Collected Papers*. Reviews by JOHN CAMPBELL, 156 (3) and SCOTT SOAMES, 141 (3).
Experience (Brute —). PETER CARRUTHERS. 258 (5).
Explanation (Reason and —) [abstract]. ANDREW LEVINE. 667 (11).

Falk, W. D. *Ought, Reasons, and Morality*. Review by STEPHEN L. DARWALL. 208 (4).
FEYERABEND, PAUL: Realism and the Historicity of Knowledge. 393 (8).
Foucault, Michel (Thought of —). APA Symposium. 601 (11).
Foucault and the Paradox of Bodily Inscriptions. JUDITH BUTLER. 601 (11).
Foucault's Critique of the Liberal Individual. DAVID F. GRUBER. 615 (11).
FRIEDMAN, MARILYN: The Impracticality of Impartiality. 645 (11).
Friedman on Impartiality and Practicality [abstract]. HOWARD MCGARY. 657 (11).

GOLDFARB, WARREN: Wittgenstein, Mind, and Scientism. 635 (11).
GRANDY, RICHARD E.: On Grice on Language. 514 (10).
Grice (Thought of —). APA Symposium. 505 (10).
GRISWOLD, CHARLES L., JR.: Adam Smith on Virtue and Self-interest [abstract]. 681 (11).
GRUBER, DAVID F.: Foucault's Critique of the Liberal Individual 615 (11).
Guyer, Paul. *Kant and the Claims of Knowlege*. Review by HENRY E. ALLISON. 214 (4).

HARRIS, GEORGE W.: A Paradoxical Departure from Consequentialism. 90 (2).
Historicity (Reason and — of Knowledge). PAUL FEYERABEND. 393 (8).
How Marxism Became Analytic. BILL MARTIN. 659 (11).
Human Agency (Social Power and —). ANDREW KERNOHAN. 712 (12).
Human Beings (Johnston on —). DAVID S. ODERBERG. 137 (3).

Ignorance (The Virtues of —). JULIA DRIVER. 373 (7).
Impartiality. APA Symposium. 645 (11).
Impracticality of Impartiality (The —). MARILYN FRIEDMAN. 645 (11).
Induction (The Backward — Paradox). PHILIP PETTIT and ROBERT SUGDEN. 169 (4).
Innovation and Understanding in Mathematics [abstract]. PHILIP KITCHER. 563 (10).
Intensionality and Perception: A Reply to Rosenberg. MOHAN MATTHEN. 727 (12).
Intentionality (Realism without Internalism: A Critique of Searle on —). AKEEL BILGRAMI. 57 (2).
Internalism (Realism without —: A Critique of Searle on Intentionality). 57 (2).

JACKSON, FRANK: Review of David Lewis: *Philosophical Papers, vol. II*. 433 (8).
Johnsonian Prize in Philosophy. 55 (1), 391 (7).
Johnston on Human Beings. DAVID S. ODERBERG. 137 (3).
JURIST, ELLIOT L. Review of Seyla Benhabib: *Critique, Norm and Utopia: A Study of the Foundations of Critical Theory*. 203 (4).
Justice (Leibniz, Keynes, and the Rabbis on a Problem of Distributive —). NICHOLAS RESCHER. 337 (7).

KADANE, JOSEPH B., TEDDY SEIDENFELD, and MARK J. SCHERVISH: On the Shared Preferences of Two Bayesian Decision Makers. 225 (5).
KERNOHAN, ANDREW: Social Power and Human Agency. 712 (12).
Keynes (Leibniz, —, and the Rabbis on a Problem of Distributive Justice). NICHOLAS RESCHER. 337 (7).
KITCHER, PHILIP: Innovation and Understanding in Mathematics [abstract]. 563 (10).
Kitcher, Philip: *Vaulting Ambition: Sociobiology and the Quest for Human Nature*. Review by RICHARD M. BURIAN. 385 (7).
Knowledge (Reason and the Historicity of —). PAUL FEYERABEND. 393 (8).
KOSSO, PETER: Science and Objectivity. 245 (5).
Kripke's Theory of Truth. APA Symposium. 530 (10).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

LARMORE, CHARLES: Review of Alasdair MacIntyre: *Whose Justice? Which Rationality?* 437 (8).
Larmore, Charles: *Patterns of Moral Complexity.* Review by JEREMY WALDRON. 331 (6).
LAYMON, RONALD: Cartwright and the Lying Laws of Physics. 353 (7).
Leibniz, Keynes, and the Rabbis on a Problem of Distributive Justice. NICHOLAS RESCHER. 337 (7).
LEVINE, ANDREW: Reason and Explanation [abstract]. 667 (11).
Lewis, Davis: *Philosophical Papers, vol. II.* Review by FRANK JACKSON. 433 (8).
Liberal Individual (Foucault's Critique of the —). DAVID F. GRUBER. 615 (11).

MACDONALD, GRAHAM: Review of Paul A. Roth: *Meaning and Method in the Social Sciences.* 442 (8).
MacIntyre, Alasdair: *Whose Justice? Which Rationality?* Review by CHARLES LARMORE. 437 (8).
MCDOWELL, JOHN: Wittgenstein and the Inner World [abstract]. 643 (11).
MCGARY, HOWARD: Friedman on Impartiality and Practicality [abstract]. 657 (11).
MCGEE, VANN: Applying Kripke's Theory of Truth. 530 (10).
MCWHORTER, LADELLE: Culture or Nature? The Function of the Term 'Body' in the Work of Michel Foucault. 608 (11).
Maladroitness of Epistemic TIT FOR TAT (The —). JOHN WOODS. 324 (6).
MANDERS, KENNETH: Domain Extension and the Philosophy of Mathematics. 553 (10).
MARQUIS, DON: Why Abortion is Immoral. 183 (4).
MARTIN, BILL: How Marxism Became Analytic. 659 (11).
Marxism (How — Became Analytic). BILL MARTIN. 659 (11).
Mathematics (The Application of — to Natural Science). MARK STEINER. 449 (9).
Mathematics (Philosophy of —). APA Symposium. 553 (10).
MATTHEN, MOHAN: Intensionality and Perception: A Reply to Rosenberg. 727 (12).
MAUDLIN, TIM: Computation and Consciousness. 407 (8).
Memo to Authors. 54 (1).
Metaphysical Construction of Value (The —). JUDITH BAKER. 505 (10).
Metapsychological Relativism and the Self. STEPHEN L. WHITE. 298 (6).
MILLIKAN, RUTH GARRETT. Biosemantics. 281 (6).
Mind (Wittgenstein, —, and Scientism). WARREN GOLDFARB. 635 (11).
Mind, Philosophy of (Wittgenstein's Later —: Sensation, Privacy, and Intention). CRISPIN WRIGHT. 622 (11).
Morality (Egoism and —). APA Symposium. 542 (10).
MULHOLLAND, LESLIE: Egoism and Morality. 542 (10).

Nature (Culture or —? The Function of the Term 'Body' in the Work of Michel Foucault. LADELLE MCWHORTER. 608 (11).
New Books. 103 (2), 277 (5).
New Books: Anthologies. 164 (3), 334 (6).
New Books: Translations. 167 (3), 446 (8).
Notes and News. 50 (1), 222 (4), 333 (6), 392 (7), 447 (8), 565 (10), 683 (11), 733 (12).

Objectivity (Science and —). PETER KOSSO. 245 (5).
Objects (Another Argument against Vague —). FRANCIS JEFFRY PELLETIER. 481 (9).
ODERBERG, DAVID S. Johnston on Human Beings. 137 (3).
On Degrees. RAYME E. ENGEL. 23 (1).
On Grandy on Grice [abstract]. ROBERT STALNAKER. 526 (10).
On Grice on Language. RICHARD E. GRANDY. 514 (10).
On the Shared Preferences of Two Bayesian Decision Makers. TEDDY SEIDENFELD, JOSEPH B. KADANE, and MARK J. SCHERVISH. 225 (5).

Paradox (Foucault and the — of Bodily Inscriptions). JUDITH BUTLER. 601 (11).
Paradox (The Backward Induction —). PHILIP PETTIT and ROBERT SUGDEN. 169 (4).
Paradox (Truth, Definite Truth, and —) [abstract]. STEVEN YABLO. 539 (10).
Paradoxical Departure from Consequentialism (A —). GEORGE W. HARRIS. 90 (2).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Perception (Intensionality and —: A Reply to Rosenberg). MOHAN MATTHEN. 727 (12).
Perceptual Presentations and Biological Function: A Comment on Matthen. Alexander Rosenberg. 38 (1).
PERRY, JOHN, and MARK CRIMMINS. The Prince and the Phone Booth: Reporting Puzzling Beliefs. 685 (12).
PETTIT, PHILLIP, and ROBERT SUGDEN. The Backward Induction Paradox. 169 (4).
Philosophy of Mathematics. APA Symposium. 553 (10).
Philosophy of Mind (Wittgenstein's Later —: Sensation, Privacy, and Intention). CRISPIN WRIGHT. 622 (11).
Power (Social — and Human Agency). ANDREW KERNOHAN. 712 (12).
Practicality (Friedman on Impartiality and —) [abstract]. HOWARD MCGARY. 657 (11).
Preference for Sense and Reference (A —). GABRIEL SEGAL. 73 (2).
Prince and the Phone Booth: Reporting Puzzling Beliefs (The —). MARK CRIMMINS and JOHN PERRY. 685 (12).
Privacy (Wittgenstein's Later Philosophy of Mind: Sensation, —, and Intention. CRISPIN WRIGHT. 622 (11).

Quine's Point of View. MIRIAM SOLOMON. 113 (3).

Realism and the Historicity of Knowledge. PAUL FEYERABEND. 393 (8).
Realism without Internalism: A Critique of Searle on Intentionality. AKEEL BILGRAMI. 57 (2).
Reason and Explanation [abstract]. ANDREW LEVINE. 667 (11).
Reference (A Preference for Sense and —). GABRIEL SEGAL. 73 (2).
Relativism (Metapsychological — and the Self). STEPHEN L. WHITE. 298 (6).
Reply to Baker and Grandy [abstract]. RICHARD WARNER. 528 (10).
RESCHER, NICHOLAS: Leibniz, Keynes, and the Rabbis on a Problem of Distributive Justice. 337 (7).
Role of Self-interest in Adam Smith's *Wealth of Nations* (The —). PATRICIA H. WERHANE. 669 (11).
ROSEN, MICHAEL E.: Review of Charles Taylor: *Philosophical Papers, 2 vols.* 270 (5).
ROSENBERG, ALEXANDER: Perceptual Presentations and Biological Function: A Comment on Matthen. 38 (1).
Roth, Paul A.: *Meaning and Method in the Social Sciences.* Review by GRAHAM MACDONALD. 442 (8).

SCHERVISH, MARK J., TEDDY SEIDENFELD, and JOSEPH B. KADANE: On the Shared Preferences of Two Bayesian Decision Makers. 225 (5).
Science and Objectivity. PETER KOSSO. 245 (5).
Science (The Application of Mathematics to Natural —). MARK STEINER. 449 (9).
Scientism (Wittgenstein, Mind, and —). WARREN GOLDFARB. 635 (11).
Searle (Realism without Internalism: A Critique of — on Intentionality). AKEEL BILGRAMI. 57 (2).
SEGAL, GABRIEL: A Preference for Sense and Reference. 73 (2).
SEIDENFELD, TEDDY, JOSEPH B. KADANE, and MARK SCHERVISH: On the Shared Preferences of Two Bayesian Decision Makers. 225 (5).
Self (Metapsychological Relativism and the —). STEPHEN L. WHITE. 298 (6).
Self-interest (Adam Smith on Virtue and —) [abstract]. Charles L. Griswold, Jr. 681 (11).
Self-interest (The Role of — in Adam Smith's *Wealth of Nations*). PATRICIA H. WERHANE. 669 (11).
Sensation (Wittgenstein's Later Philosophy of Mind: —, Privacy, and Intention). CRISPIN WRIGHT. 622 (11).
Sense (A Preference for — and Reference). GABRIEL SEGAL. 73 (2).
Smith, Adam (Thought of —). APA Symposium. 669 (11).
SOAMES, SCOTT: Review of Gareth Evans: *Collected Papers.* 141 (3).
Social Power and Human Agency. ANDREW KERNOHAN. 712 (12).
SOLOMON, MIRIAM: Quine's Point of View. 113 (3).
STALNAKER, ROBERT: On Grandy on Grice [abstract]. 526 (10).
STEINER, MARK: The Application of Mathematics to National Science. 449 (9).
SUGDEN, ROBERT, and PHILLIP PETTIT. The Backward Induction Paradox. 169 (4).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Taylor, Charles: *Philosophical Papers, 2 vols.* Review by MICHAEL E. ROSEN.
THOMAS, LAURENCE: Doing Justice to Egoism [abstract]. 551 (10).
Thought of Grice. APA Symposium. 505 (10).
Thought of Adam Smith. APA Symposium. 669 (11).
Truth (Kripke's Theory of —). APA Symposium. 530 (10).
Truth, Definite Truth, and Paradox [abstract]. STEPHEN YABLO. 539 (10).

Vague Objects (Another Argument against —). FRANCIS JEFFRY PELLETIER. 481 (9).
Value (On the Metaphysical Construction of —). JUDITH BAKER. 505 (10).
Virtue (Adam Smith on — and Self-interest). CHARLES L. GRISWOLD, JR. 681 (11).
Virtues of Ignorance (The —). JULIA DRIVER. 373 (7).

WALDRON, JEREMY: Review of Charles Larmore: *Patterns of Moral Complexity.* 331 (6).
WARNER, RICHARD: Reply to Baker and Grandy [abstract]. 528 (10).
WERHANE, PATRICIA H.: The Role of Self-interest in Adam Smith's *Wealth of Nations.* 669 (11).
WHITE, STEPHEN L. Metapsychological Relativism and the Self. 298 (6).
Why Abortion is Immoral. DON MARQUIS. 183 (4).
Why Dialogue? BRUCE ACKERMAN. 5 (1).
Wittgenstein and the Inner World [abstract]. JOHN MCDOWELL. 643 (11).
Wittgenstein's Later Philosophy of Mind: Sensation, Privacy, and Intention. CRISPIN WRIGHT. 622 (11).
Wittgenstein, Mind, and Scientism [abstract]. WARREN GOLDFARB. 635 (11).
Wittgenstein (Thought of —). APA Symposium. 622 (11).
WOODS, JOHN. The Maladroitness of Epistemic TIT FOR TAT. 324 (6).
WRIGHT, CRISPIN: Wittgenstein's Later Philosophy of Mind: Sensation, Privacy, and Intention. 622 (11).

YABLO, STEPHEN: Truth, Definite Truth, and Paradox [abstract]. 539 (10).

"The subtlety and skill, the range, depth and power of the arguments deployed in Schiffer's book are such as will be expected by any philosopher acquainted with his previous work. That is to say they are truly impressive . . . Schiffer's book is probably the most significant work in the field that has appeared for a good many years." — P. F. Strawson, Times Literary Supplement

Now Available in Paperback

REMNANTS OF MEANING

STEPHEN SCHIFFER

Schiffer surveys all the leading theories of meaning and content in the philosophy of language and concludes that there can be no correct, positive philosophical theory of linguistic or mental representation and, accordingly, advocates the deflationary "no-theory theory of meaning and content." Along the way, he takes up functionalism, the nature of propositions and their suitability as contents, the language of thought and other sententialist theories of belief, intention-based semantics, and related issues in ontology.

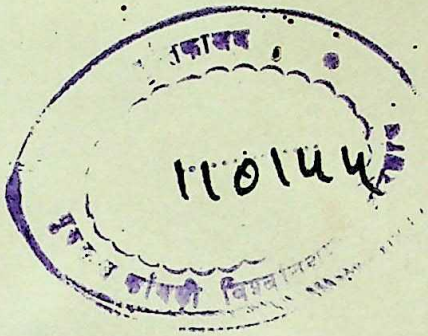
"There are only perhaps a half-dozen philosophers in the world who can address these issues with Schiffer's authority and depth of insight."
— Stephen P. Stich, Rutgers University

A Bradford Book. 303 pp. $32.50 cloth, $12.95 paper

55 Hayward Street **The MIT Press** Cambridge, MA 02142

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri



CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

110144

Digitized by Arya Samaj Foundation Chennai and eGangotri

Compiled
1999-2000

Ambika Book Binder
Bill No./date

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar